Jurgen A. Doornik and David F. Hendry

Econometric Modelling

PcGive™ 14 Volume III

OxMetrics™

Published by Timberlake Consultants Ltd
www.timberlake.co.uk
www.timberlake-consultancy.com
# Contents

## Front matter
- iii

## Contents
- v

## List of Figures
- ix

## List of Tables
- xi

## Preface
- xiii

## I Prologue

1 Introduction to Volume III

1.1 The PcGive system ........................................ 3
1.2 Citation .................................................... 4
1.3 World Wide Web ........................................... 4

## II Limited Dependent Models (LogitJD)

2 Discrete choice models

2.1 Introduction .................................................. 7
2.2 Binary discrete choice ...................................... 7
2.3 The binary logit and probit model ....................... 8
2.4 Multinomial discrete choice .............................. 8
  2.4.1 The multinomial logit model .......................... 10
  2.4.2 Weighted estimation .................................. 10
2.5 Evaluation .................................................. 11
  2.5.1 Estimated probabilities .............................. 11
  2.5.2 Likelihood ratio tests ................................ 11
  2.5.3 Derivatives of probabilities ......................... 12
2.6 Histograms ................................................ 13
2.7 Norm observations ........................................ 13
## CONTENTS

2.8 Observed versus predicted ........................................ 14
2.9 Outlier analysis ....................................................... 14

3 Tutorial on Discrete Choice Modelling ........................... 15
3.1 Introduction .......................................................... 15
3.2 Data organization ...................................................... 16
3.3 Binary logit estimation .............................................. 17
3.4 Binary probit estimation ............................................ 20
3.5 Grouped logit estimation ............................................ 21
3.6 Multinomial logit estimation ....................................... 22
3.7 Conditional logit estimation ....................................... 24
3.8 Automatic model selection for binary logit and probit ....... 27

III Panel Data Models (DPD)
* with Manuel Arellano and Stephen Bond* ............................... 29

4 Panel Data Models ......................................................... 31
4.1 Introduction .......................................................... 31
4.2 Econometric methods for static panel data models .......... 32
   4.2.1 Static panel-data estimation ................................ 32
4.3 Econometric methods for dynamic panel data models ....... 33

5 Tutorial on Static Panel Data Modelling ............................. 40
5.1 Introduction .......................................................... 40
5.2 Data organization ...................................................... 40
5.3 Static panel data estimation ....................................... 42

6 Tutorial on Dynamic Panel Data Modelling .......................... 47
6.1 Introduction .......................................................... 47
6.2 Data organization ...................................................... 47
6.3 One-step GMM estimation ........................................... 48
6.4 Two-step GMM estimation ........................................... 52
6.5 IV estimation .......................................................... 53
6.6 Combined GMM estimation ......................................... 54

7 Panel Data Implementation Details .................................... 57
7.1 Transformations ....................................................... 57
7.2 Static panel-data estimation ....................................... 58
7.3 Dynamic panel data estimation .................................... 59
7.4 Dynamic panel data, combined estimation ....................... 63
7.5 Panel batch commands .............................................. 64
IV Volatility Models (GARCH)
with H. Peter Boswijk and Marius Ooms 67

8 Introduction to Volatility Models 69
  8.1 Introduction 69

9 Tutorial on GARCH Modelling 73
  9.1 Estimating a GARCH(1, 1) model 73
  9.2 Evaluating the GARCH(1, 1) model 77
  9.3 Recursive estimation of the GARCH(1, 1) model 79
  9.4 GARCH(1, 1) with regressors in the variance equation 82
  9.5 GARCH(1, 1) with Student t-distributed errors 85
  9.6 EGARCH(1, 1) GED-distributed errors 86
  9.7 GARCH in mean 87
  9.8 Asymmetric threshold GARCH 90
  9.9 GARCH batch usage 92

10 GARCH Implementation Details 94
  10.1 GARCH model settings 94
  10.2 Some implementation details 96
  10.3 GARCH batch commands 97

V Time Series Models (ARFIMA)
with Marius Ooms 99

11 Introduction to Time Series Models 101

12 Tutorial on ARFIMA Modelling 103

13 ARFIMA Implementation Details 115
  13.1 Introduction 115
  13.2 The Arfima model 115
    13.2.1 Autocovariance function 116
  13.3 Estimation 117
    13.3.1 Regressors in mean 117
    13.3.2 Initial values 117
    13.3.3 Exact maximum likelihood (EML) 118
    13.3.4 Modified profile likelihood (MPL) 119
    13.3.5 Non-linear least squares (NLS) 119
    13.3.6 Variance-covariance matrix estimates 119
  13.4 Estimation output 120
  13.5 Estimation options 120
13.5.1 Sample mean versus known mean ............................. 120
13.5.2 Fixing parameters ............................................. 121
13.5.3 Weighted estimation ........................................... 121
13.5.4 $Z$ variables .................................................. 121
13.6 Forecasting ....................................................... 122
13.7 ARFIMA batch commands ....................................... 123

### VI X12arima for OxMetrics

#### 14 Overview of X12arima for OxMetrics

14.1 Introduction .................................................. 129
14.2 X-12-ARIMA .................................................... 129
14.3 Credits ......................................................... 130
14.4 Disclaimer ..................................................... 130
14.5 Limitations ..................................................... 130
14.6 Documentation ................................................ 130
14.7 Census X-11 Seasonal Adjustment ............................. 131
14.8 X-12-ARIMA Seasonal Adjustment ............................ 132
14.9 regARIMA ....................................................... 132
14.10 ARIMA model selection ....................................... 133
14.11 X12arima in PcGive .......................................... 134

#### 15 Tutorial on Seasonal Adjustment with X12arima for OxMetrics

15.1 Introduction .................................................. 136
  15.1.1 Loading the first data set ................................. 136
  15.1.2 Quick Seasonal Adjustment ............................... 136
  15.1.3 Test menu .................................................. 143

#### 16 Tutorial on regARIMA Modelling with X12arima for OxMetrics

16.1 Introduction .................................................. 145
  16.2 Estimating the Airline model ............................... 145
    16.2.1 Model Output .......................................... 146
  16.3 regARIMA Model Example ................................. 148
    16.3.1 Model Output .......................................... 149

References ......................................................... 153

Author Index ....................................................... 159

Subject Index ...................................................... 161
# Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Example histograms of probabilities</td>
<td>13</td>
</tr>
<tr>
<td>3.1</td>
<td>Cramer(2003) car-ownership data</td>
<td>16</td>
</tr>
<tr>
<td>3.2</td>
<td>OLS estimation of CAR01</td>
<td>17</td>
</tr>
<tr>
<td>3.3</td>
<td>Probabilities at sample mean</td>
<td>24</td>
</tr>
<tr>
<td>3.4</td>
<td>Graphic Analysis: histogram of probabilities</td>
<td>25</td>
</tr>
<tr>
<td>5.1</td>
<td>Grunfeld(1958) investment data</td>
<td>41</td>
</tr>
<tr>
<td>5.2</td>
<td>Graphical output from four static panel estimates</td>
<td>45</td>
</tr>
<tr>
<td>9.1</td>
<td>Daily returns on DM/UK£ exchange rate, ACF and ACF of squares</td>
<td>74</td>
</tr>
<tr>
<td>9.2</td>
<td>GARCH(1, 1) model for daily returns on DM/UK£ exchange rate, residual ACF and ACF of squared residuals</td>
<td>78</td>
</tr>
<tr>
<td>9.3</td>
<td>Estimated density of scaled residuals of GARCH(1, 1) model</td>
<td>79</td>
</tr>
<tr>
<td>9.4</td>
<td>Recursively estimated GARCH(1, 1) model</td>
<td>80</td>
</tr>
<tr>
<td>9.5</td>
<td>Recursively estimated GARCH(1, 1) model with and without $\alpha_1 + \beta_1 &lt; 1$</td>
<td>81</td>
</tr>
<tr>
<td>9.6</td>
<td>Forecasts from GARCH(1, 1) model with and without trading day effect</td>
<td>84</td>
</tr>
<tr>
<td>9.7</td>
<td>Forecasts from GARCH(1, 1) and GARCH-t(1, 1) models</td>
<td>85</td>
</tr>
<tr>
<td>9.8</td>
<td>Graphical analysis for GARCH-M(1, 1) model</td>
<td>88</td>
</tr>
<tr>
<td>9.9</td>
<td>Recursive graphics for GARCH-M(1, 1) model</td>
<td>89</td>
</tr>
<tr>
<td>9.10</td>
<td>Recursive graphics for unrestricted GARCH-M(1, 1) model</td>
<td>90</td>
</tr>
<tr>
<td>9.11</td>
<td>Forecasts from restricted and unrestricted GARCH-M(1, 1) model</td>
<td>90</td>
</tr>
<tr>
<td>9.12</td>
<td>Recursive graphics for TAGARCH(1, 1) model</td>
<td>91</td>
</tr>
<tr>
<td>9.13</td>
<td>Forecasts for TAGARCH(1, 1) model</td>
<td>92</td>
</tr>
<tr>
<td>12.1</td>
<td>Quarterly and annual inflation in the UK, based on RPI (all items)</td>
<td>104</td>
</tr>
<tr>
<td>12.2</td>
<td>Graphical analysis of seasonal ARFIMA$(4, d, 4)$ model for UK inflation</td>
<td>110</td>
</tr>
<tr>
<td>12.3</td>
<td>Forecasts from ARFIMA$(4, d, 4)$ model for UK inflation</td>
<td>111</td>
</tr>
<tr>
<td>12.4</td>
<td>Forecasts of annual changes from ARFIMA$(4, d, 4)$ model for UK inflation</td>
<td>112</td>
</tr>
<tr>
<td>Number</td>
<td>Figure Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>12.5</td>
<td>Levels forecasts from ARFIMA$(4, d, 4)$ model for UK inflation</td>
<td>113</td>
</tr>
<tr>
<td>15.1</td>
<td>Time series graph of ICMETI</td>
<td>137</td>
</tr>
<tr>
<td>15.2</td>
<td>Seasonally adjusted ICMETI</td>
<td>139</td>
</tr>
<tr>
<td>15.3</td>
<td>Spectral plots for ICMETI X11 decomposition</td>
<td>144</td>
</tr>
<tr>
<td>16.1</td>
<td>Graphic analysis of airline model</td>
<td>147</td>
</tr>
<tr>
<td>16.2</td>
<td>Seasonal adjustment of BeAuto series</td>
<td>150</td>
</tr>
</tbody>
</table>
Tables

4.1 Static and dynamic panel data estimators of PcGive . . . . . . . 31
5.1 Static panel estimates for Grunfeld data . . . . . . . . . . . . . 46
7.1 Transformation and the treatment of dummy variables . . . . . . . 60
14.1 Summary of steps taken in the Census X-11 method . . . . . . . 131
Preface

This book is the third volume of the core PcGive documentation. Volume I and II focus on univariate and multivariate dynamic econometric modelling respectively. This volume describes a wide range of econometric techniques, ranging from ARFIMA, Markov-switching and GARCH modelling through logit and panel data to X-12-ARIMA seasonal adjustments. Most of these have already been tested for several years as Ox packages, and we wish to thank the many users who commented on the earlier versions. X-12-ARIMA is created by the US Census Bureau; the OxMetrics version is based on the FORTRAN code which they make publicly available — this has been a long ongoing research project and we wish to thank all the involved people for sharing their output.

This book is different from the other two volumes in that we have had significant help from our co-authors, Manuel Arellano, Stephen Bond, Peter Boswijk and Marius Ooms. We are grateful to them for offering their support. We also wish to thank Statistics Netherlands to allow us to use the car ownership data set for logit modelling.

MikTeX in combination with dvips and GhostView eased the development of the documentation in \LaTeX. The editing was almost entirely undertaken using OxEdit, which allowed flexible incorporation of PcGive output and LaTeX compilation, and HyperSnap-DX which enabled the screen captures (with bmp2eps for conversion to PostScript).

The authors owe a considerable debt to their families for the long hours they spent on this project: their support and encouragement were essential, even though they could only benefit indirectly from the end product. We hope the benefits derived by others compensate.

We wish you enjoyable and productive use of

PcGive for Windows
Part I

Prologue
Chapter 1

Introduction to Volume III

1.1 The PcGive system


The econometric techniques of the PcGive system can be divided by the type of data to which they are (usually) applied. The documentation is comprised of three volumes, and the overview below gives in parenthesis whether the method is described in Volume I, II or III. Volume IV refers to the PcNaive book.

- Models for cross-section data
  - Cross-section Regression (I)
- Models for discrete data
  - Binary Discrete Choice (III): Logit and Probit
  - Multinomial Discrete Choice (III): Multinomial Logit
  - Count data (III): Poisson and Negative Binomial
- Models for financial data
  - GARCH Models (III): GARCH in mean, GARCH with Student-t, EGARCH, Estimation with Nelson&Cao restrictions
- Models for panel data
  - Static Panel Methods (III): within groups, between groups
  - Dynamic Panel Methods (III): Arellano-Bond GMM estimators
- Models for time-series data
  - Single-equation Dynamic Modelling (I)
  - Multiple-equation Dynamic Modelling (II): VAR and cointegration, simultaneous equations analysis
  - Regime Switching Models (III): Markov-switching
  - ARFIMA Models (III): exact maximum likelihood, modified-profile likelihood
or non-linear least squares

- **Monte Carlo**
  - AR(1) Experiment using PcNaive (IV)
  - Static Experiment using PcNaive (IV)
  - Advanced Experiment using PcNaive & Ox Professional (IV)

- **Other models**
  - Nonlinear Modelling (I)
  - Descriptive Statistics (I)

The current book, **Volume III**, describes the entries marked with III above. PcGive uses OxMetrics for data input and graphical and text output. OxMetrics is described in a separate book (Doornik and Hendry, 2013). Even though PcGive is largely written in Ox (Doornik, 2013), it does not require Ox to function.

### 1.2 Citation

To facilitate replication and validation of empirical findings, PcGive should be cited in all reports and publications involving its application. The appropriate form is to cite PcGive in the list of references.

### 1.3 World Wide Web

Consult [www.oxmetrics.net](http://www.oxmetrics.net) or [www.doornik.com](http://www.doornik.com) pointers to additional information relevant to the current and future versions of PcGive. Upgrades are made available for downloading if required, and a demonstration version is also made available.
Part II

Limited Dependent Models
(LogitJD)
Chapter 2

Discrete choice models

2.1 Introduction

In a discrete choice model the dependent variable only takes on integer values. Such a
discrete dependent variable can denote a category (e.g. mode of transport: car, train, or
bus) or count the number of events (e.g. the number of accidents in a week). A model
with a categorical dependent variable is often called a discrete choice model, but is also
referred to as dummy-endogenous or qualitative response model. As our main reference
for discrete choice models we use Cramer (2003).

The use of ordinary least squares in a discrete choice model is called the linear prob-
ability model. The main shortcoming of this model is that it can predict probabilities
that are outside the [0, 1] interval. PcGive implements logit and probit estimation which
avoids this drawback.

General references include McFadden (1984), Amemiya (1981) and Cramer (2003),
among others. In the remainder we exclusively refer to Cramer (2003).

2.2 Binary discrete choice

In the simple discrete choice model the dependent variable only takes on two values,
for example, when modelling car ownership:

\[ y_i = 1 \quad \text{if household } i \text{ owns a private car,} \]
\[ y_i = 0 \quad \text{otherwise.} \]

With a discrete dependent variable, interest lies in modelling the probabilities of ob-
serving a certain outcome:

\[ p_i = P \{ y_i = 1 \} . \]

Using OLS on a dummy dependent variable:

\[ y_i = x_i' \beta + \varepsilon_i \]
does not restrict \( \hat{y}_i \) between 0 and 1, which would be required if it is to be interpreted as a probability. Also, the disturbances cannot be normally distributed, as they only take on two values: \( \varepsilon_i = 1 - p_i \) or \( \varepsilon_i = 0 - p_i \), writing \( p_i = x_i' \beta \), and are heteroscedastic.

A simple solution is to introduce an underlying continuous variable \( U_i \), which is not observed. Observed is:

\[
y_i = \begin{cases} 
0 & \text{if } U_i < 0, \\
1 & \text{if } U_i \geq 0.
\end{cases}
\]  

(2.1)

Now we can introduce explanatory variables:

\[
U_i = x_i' \beta - \varepsilon_i.
\]

and write

\[
p_i = P \{ y_i = 1 \} = P \{ x_i' \beta - \varepsilon_i \geq 0 \} = F_{\varepsilon} (x_i' \beta).
\]

Observations with \( y_i = 1 \) contribute \( p_i \) to the likelihood, observations with \( y_i = 0 \) contribute \( 1 - p_i \):

\[
L (\beta \mid X) = \prod_{\{y_i = 0\}} (1 - p_i) \prod_{\{y_i = 1\}} p_i,
\]

and the log-likelihood becomes:

\[
\ell (\beta \mid X) = \sum_{i=1}^{N} [(1 - y_i) \log (1 - p_i) + y_i \log p_i].
\]

### 2.3 The binary logit and probit model

The choice of \( F_{\varepsilon} \) determines the method. Using the logistic distribution:

\[
F_{\varepsilon} (z) = \frac{e^z}{1 + e^z}
\]

leads to logit. Logit has a linear log-odds ratio:

\[
\log \left( \frac{p_i}{1 - p_i} \right) = x_i' \beta.
\]

The standard normal distribution gives probit. We can multiply \( y_i^* \) by any non-zero constant without changing the outcome, the scale of these distributions is fixed: the logistic has variance \( \pi^2 / 3 \), the standard normal has variance equal to 1.

### 2.4 Multinomial discrete choice

Assume that observations on \( n \) individuals (or households, firms, etc.) are obtained by random sampling, and that all individuals have made a choice from the same set of alternatives \( A = \{1, 2, \ldots, S\} \). Using subscript \( i \) to denote an individual and \( s \) to denote an alternative, we can write the dependent variable as \( y_i \):

\[
y_i = s \quad \text{if individual } i \text{ has made choice } s, \quad s = 1, \ldots, S; \quad i = 1, \ldots, n.
\]
The existence of an unobserved continuous variable $U_{is}$ is assumed, where the alternative with the highest value is chosen:

$$y_i = s \text{ if } U_{is} = \max_{k \in A} \{U_{ik}\}.$$  

See Cramer (2003, Ch. 2) for an interpretation in terms of utility maximization. Adding a stochastic component:

$$U_{is} = V_{is} + \varepsilon_{is}. \tag{2.2}$$

Logit and probit specify a certain distribution for $\varepsilon_{is}$. The systematic part $V_{is}$ is a functional form of explanatory variables and coefficients. We can distinguish three types of explanatory variables:

1. depending only on the characteristics of individual $i$ (denoted $x_i$);
2. depending only on the alternative $s$ (denoted $r_s$);
3. depending on both (denoted $z_{is}$).

The $x_i$ are called alternative independent regressors, while the remaining two types, called alternative dependent, are treated identically in PcGive.

Moreover, there are two types of coefficients:

1. alternative specific (denoted $\beta_j$);
2. generic (denoted $\gamma$).

The full specification in PcGive is:

$$V_{is} = x_i' \beta_s + z_{is}' \gamma. \tag{2.3}$$

So alternative independent regressors always enter with an alternative specific coefficient, while alternative dependent regressors have a generic coefficient.

As an illustration, consider the transport mode example:

- $y_i = 1$ if mode of transport is car,
- $y_i = 2$ if mode of transport is bus,
- $y_i = 3$ if mode of transport is train.

Then $x_i$ may be the income of individual $i$, and $z_{is}$ the travel time for individual $i$ using mode $s$. Interest may focus on the probabilities that individual $i$ uses mode $s$, or the coefficients, to estimate the influence on marginal utility of explanatory variables. In the absence of alternative specific parameters it is possible to investigate the consequences of introducing a new mode of transport.

Let $p_{is}$ denote the probability that individual $i$ chooses alternative $s$. Then from (2.2):

$$p_{is} = \Pr(y_i = s) = \Pr(U_{is} = \max_{k \in A} \{U_{ik}\}) = \Pr(V_{is} + \varepsilon_{is} > V_{ik} + \varepsilon_{ik} \text{ for all } k \neq s).$$

The log-likelihood function is:

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \sum_{s=1}^{S} Y_{is} \log P_{is}(\theta), \tag{2.4}$$
where
\[
Y_{is} = \begin{cases} 
1 & \text{if } y_i = s, \\
0 & \text{otherwise.}
\end{cases}
\]

### 2.4.1 The multinomial logit model

The multinomial logit (MNL) model arises from (2.2) when the \( \varepsilon_{ik} \) are independently type I extreme value distributed (see Cramer, 2003, Ch. 3). This is the most convenient model, because of the closed form of the probabilities:

\[
p_{is} = \frac{\exp (V_{is})}{1 + \sum_{k=1, k \neq s}^S \exp (V_{ik} - V_{is})},
\]

and the concavity of the log-likelihood.

The log-odds ratio is:

\[
\log \left( \frac{p_{is}}{p_{ik}} \right) = V_{is} - V_{ik}.
\]

Substituting (2.3):

\[
V_{is} - V_{ik} = x_i' (\beta_s - \beta_k) + (z_{is} - z_{ik})' \gamma,
\]

shows that a normalization is needed on \( \beta_k \): a constant can be added (for all \( k \)) without changing the probabilities. PcGive takes:

\[
\beta_1 = 0,
\]

and the first alternative is then called the reference state.

Although the name multinomial logit in PcGive is generally used for the specification (2.3), it sometimes stands for a more specific functional form:

\[
V_{is} = x_i' \beta_s.
\]

In that case, the conditional logit model refers to

\[
V_{is} = z_{is}' \gamma.
\]

When \( V_{is} \) does not depend on attributes of alternatives other than alternative \( s \), the MNL model has the independence from irrelevant alternatives (IIA) property.

### 2.4.2 Weighted estimation

PcGive allows the specification of a weight variable in the log-likelihood:

\[
\ell(\theta) = \sum_{i=1}^n w_i \sum_{s=1}^S Y_{is} \log P_{is}(\theta).
\]

The use could be two-fold:
1. Weighted M-estimation (Ruud, 1986)
   To get consistent estimates of the slope parameters (up to an unknown scale factor) in the case of a mis-specified distribution (e.g. using logit when the true model is probit, although in the binary case the models are very close). In discrete choice models weighting is not always necessary, see Ruud (1983).

2. Choice-based sampling
   In multinomial logit, only the intercept is incorrect when treating a choice-based sample as a random sample (see Maddala, 1983, pp 90-91). Manski and Lerman (1977) proposed a weighted estimation to get consistent estimates for choice-based samples. Let \( Q_0(s) \) be the population fraction choosing alternative \( s \), and \( H(s) \) the fraction in the sample choosing \( s \). Denote the observed response of individual \( i \) as \( s_i \), then the weights are:
   \[
   w_i = \frac{Q_0(s_i)}{H(s_i)}.
   \]
   Also see Amemiya and Vuong (1987).

2.5 Evaluation

2.5.1 Estimated probabilities
   The coefficient estimates \( \hat{\beta}_s, s = 2, \ldots, S \) and \( \hat{\gamma} \) form the basis for the estimated probabilities for each observation: \( \hat{p}_{is}, s = 1, \ldots, S, i = 1, \ldots, n \).

2.5.2 Likelihood ratio tests
   If the estimated model contains a constant term, it can be compared with the baseline model in a likelihood ratio test. The baseline model has only an intercept, and the resulting test is analogous to the regression F-test. In the baseline MNL model:
   \[
   V_{is} = \alpha_s, \quad s = 2, \ldots, S,
   \]
   the likelihood can be maximized explicitly, resulting in:
   \[
   \alpha_s = \log \left( \frac{f_s}{f_1} \right),
   \]
   where \( f_s \) is the relative frequency in the sample of alternative \( s \). With \( K \) alternative independent regressors \( (x) \) and \( C \) alternative dependent regressors \( (z) \), the estimated model has \( K(S - 1) + C \) parameters, and a log-likelihood of \( \ell(\hat{\theta}) \), say. The baseline model has \( S - 1 \) parameters, and a log-likelihood of \( \ell(\hat{\alpha}) \), so that \( 2\ell(\hat{\theta}) - 2\ell(\hat{\alpha}) \) is \( \chi^2 \left[ (K - 1)(S - 1) + C \right] \) distributed.
   If the estimated model has no intercept, the zeroline is used in the test. This is a model without parameters:
   \[
   \hat{p}_{is} = \frac{1}{S}.
   \]
In a multinomial logit model that does have an intercept, the sum of the estimated probabilities for state $s$ equals the number of observations having state $s$: $\frac{1}{n} \sum_{i=1}^{n} \hat{p}_{is} = f_s$. This follows, because at the maximum, the derivatives of the log-likelihood:

$$\frac{\partial \ell}{\partial \beta_s} = \sum_{i=1}^{n} (Y_{is} - p_{is}) x_i'$$

are zero, and hence, if one of the regressors is a row of ones:

$$\sum_{i=1}^{n} Y_{is} = \sum_{i=1}^{n} \hat{p}_{is}.$$ 

With grouped data, the log-likelihood of the saturated model is reported. This is a model with perfect fit: the predicted frequencies equal the observed frequencies. Let $n_{is}$ be the number in group $i$ having response $s$. Then $\sum_{s=1}^{S} n_{is} = n_i$ and $\sum_{i=1}^{G} n_{i} = n$, with $G$ the number of groups and $n$ the total number of observations. The saturated model

$$\hat{p}_{is} = \frac{n_{is}}{n_i}$$

has $G(1 - S)$ parameters and a log-likelihood of

$$\sum_{i=1}^{G} \sum_{s=1}^{S} n_{is} \log \frac{n_{is}}{n_i}.$$ 

The saturated log-likelihood is not reported if any cell has $n_{is} = 0$.

### 2.5.3 Derivatives of probabilities

The derivatives of the probabilities with respect to the regressors give the effect of a change in one of the regressors on the probabilities. Since the probabilities sum to one, the derivatives sum to zero. The derivatives for the MNL model are:

$$\frac{\partial p_s}{\partial x_m} = \hat{p}_s (\beta_{sm} - \sum_{k=1}^{S} \hat{\beta}_{km} \hat{p}_k),$$

$$\frac{\partial p_s}{\partial z_{sm}} = \gamma_m \hat{p}_s (1 - \hat{p}_s),$$

$$\frac{\partial p_s}{\partial z_{rm}} = \gamma_m \hat{p}_s \hat{p}_r, \quad r \neq s.$$ 

We can substitute the relative frequency of each state for the probabilities ($\hat{p}_s = f_s$) or select other relative frequencies. In the calculation of the $t$-values the $\hat{p}_s$ are treated as non-stochastic.

In the binary probit model the derivatives are:

$$\frac{\partial p_s}{\partial x_m} = \phi \left( \hat{V} \right) \beta_m.$$ 

These derivatives are calculated at the mean of the explanatory variables ($\hat{V} = x' \hat{\beta}$), but can be calculated at other values.
PcGive also reports quasi-elasticities:

\[
\frac{\partial p_s}{\partial \log x_m} = \frac{\partial p_s}{\partial x_m} \bar{x}_m.
\]

If \( x_m \) increases with \( q\% \), \( p_s \) will approximately increase by \( q \times \frac{\partial p_s}{\partial \log x_m} \) percentage points. Elasticities are also reported:

\[
\frac{\partial p_s}{\partial x_m} \frac{x_m}{p_s}.
\]

All three measures have the same \( t \)-values.

## 2.6 Histograms

The distribution of the estimated probabilities is shown in histograms. The first histogram is the frequency distribution of the probabilities of the observed state. Consider the following example with three states:

<table>
<thead>
<tr>
<th>observed state</th>
<th>( \hat{p}_1 )</th>
<th>( \hat{p}_2 )</th>
<th>( \hat{p}_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>0.8</td>
<td>0.1</td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
</tr>
</tbody>
</table>

The probabilities of the observed state are underlined. Their frequency distribution is shown in Figure 2.1a (e.g. 50% is in \([0.4, 0.5]\)), while on the right is the distribution of the estimated probabilities for the first state. These graphs can be made for each state, with \( S + 1 \) histograms in total.

![Figure 2.1 Example histograms of probabilities](image)

## 2.7 Norm observations

A norm observation has the sample mean for each explanatory variable. The probabilities of each response can be calculated for the norm observation, as well as for other
user-defined norm observations. The probabilities of the norm observation do not equal the relative frequencies of the alternatives: $\hat{p}_s(\bar{V}) \neq f_s$.

2.8 Observed versus predicted

The predicted choice is the alternative with the highest predicted probability. The table of observed versus predicted choice is a rather crude measure, since 0.34, 0.33, 0.33 and 0.9, 0.05, 0.05 both select the first alternative as predicted response. For the example of §2.6 the table would read:

<table>
<thead>
<tr>
<th>Observed response</th>
<th>Predicted State 1</th>
<th>Predicted State 2</th>
<th>Predicted State 3</th>
<th>Predicted total</th>
</tr>
</thead>
<tbody>
<tr>
<td>State 1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>State 2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>State 3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Observed total</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

2.9 Outlier analysis

Let $\hat{p}(i)$ denote the estimated probability that individual $i$ selects the alternative she actually choose. These are underlined in the example of §2.6. PcGive can print all $\hat{p}(i)$ which are smaller than a designated value.
Chapter 3

Tutorial on Discrete Choice Modelling

3.1 Introduction

The examples for discrete choice modelling are based on Cramer (2003), to which we refer for further details.

Cramer (2003, §§3.6,3.7) studies the determinants of the ownership of private cars from the Expenditure Survey 1980 by Statistics Netherlands.¹ This survey provides extensive and detailed information about income and expenditure of a sample of 2820 households.

The data set is provided as NLcar.in7/NLcar.bn7, and contains:

- CAR, the number of cars available to the household, encoded as follows:
  0: NONE, household does not own private car (1010, or 36%);
  1: USED, household owns one used private car (944, or 33%);
  2: NEW, household owns one new private car (691, or 25%);
  3: MORE, household owns more than one private car (175, or 6%).
- INC, household income per equivalent adult (in Dutch guilders);
- SIZE, the household size, measured by the number of equivalent adults;
- BUSCAR, a (0, 1) dummy variable for the presence of a business car in the household;
- URBA, the degree of urbanization, measured on a six-point scale from the countryside (1) to the city (6);
- HHAGE, the age of the head of household, measured by five-year classes, starting with the class ‘below 20.’
- LINC, log(INC);
- LSIZE, log(SIZE);
- CAR01, a (0, 1) dummy variable for the presence of a car in the household.

¹This data set has been kindly made available for study purposes by Statistics Netherlands, www.cbs.nl/en.
3.2 Data organization

For multinomial logit, the dependent variable can be organized in two ways:

1. Single variable which holds the observed choice. For example, with $S$ choices, this variable takes on the values $1, 2, \ldots, S$. PcGive, will automatically make the necessary adjustments if the states are not numbered consecutively, or from another base.

2. One variable for each category. For example, $y_1, y_2, \ldots, y_S$, where each only takes on the values 0 or 1.

3. One variable for each category, but the $y_s$ are not (0, 1) dummy variables: in this case, grouped logit or probit is estimated.

Figure 3.1a shows a scatter plot of CAR versus LINC. Because CAR only takes on the values 0, 1, 2, 3, it is difficult to interpret the graph. In the second graph we have randomized CAR by adding $0.5$ times a uniform random number.

![Figure 3.1](image)

Cramer(2003) car-ownership data

OLS estimation of CAR01 on a constant and LINC can be done with the Cross-section Regression using PcGive model class in OxMetrics:

EQ( 1) Modelling CAR01 by OLS-CS

The estimation sample is: 1 - 2820

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
<th>Part.R$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.122589</td>
<td>0.1874</td>
<td>-0.654</td>
<td>0.513</td>
</tr>
<tr>
<td>LINC</td>
<td>0.0790938</td>
<td>0.01936</td>
<td>4.08</td>
<td>0.000</td>
</tr>
</tbody>
</table>

sigma 0.478215 RSS 644.446381
R$^2$ 0.00588655 F(1,2818) = 16.69 [0.000]**
3.3 Binary logit estimation

We now assume that you have loaded the NLcar.in7 data set into OxMetrics and have started PcGive. In OxMetrics, select Model, Models for discrete data and then the Binary Discrete Choice model class:
Next, select Formulate, and formulate the model with *CAR01* as dependent variable and *LINC* as regressor. A constant is added automatically. The model now is:

![Image of Formulate - Binary Discrete Choice - NLcar.in7](image)

Click on OK, to see:

![Image of Model Settings - Binary Discrete Choice](image)

The default is a logit model, so keep that, and press OK again:

![Image of Estimate - Binary Discrete Choice](image)

Again keep of the default of using Newton’s method for maximizing the log-likelihood. The results appear quickly:

```
CS( 1) Modelling CAR01 by LOGIT
  The dataset is: .\OxMetrics7\data\NLcar.in7
  The estimation sample is 1 - 2820

  Coefficient Std.Error t-value t-prob
  Constant     -2.77231  0.8283  -3.35  0.001
  LINC         0.347582  0.08579  4.05  0.000

  log-likelihood -1831.286
  no. of states  2
  no. of observations  2820
  baseline log-lik -1839.627
  Test: Chi^2( 1) 16.681 [0.0000]**
  AIC 3666.572
  AIC/T 1.30020283
  mean(CAR01) 0.641844
  var(CAR01) 0.22988
  Newton estimation (eps1=0.0001; eps2=0.005): Strong convergence
```
### 3.3 Binary logit estimation

<table>
<thead>
<tr>
<th>State</th>
<th>Count</th>
<th>Frequency</th>
<th>Probability</th>
<th>loglik</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1010</td>
<td>0.35816</td>
<td>0.35816</td>
<td>-1032.</td>
</tr>
<tr>
<td>1</td>
<td>1810</td>
<td>0.64184</td>
<td>0.64184</td>
<td>-799.3</td>
</tr>
<tr>
<td>Total</td>
<td>2820</td>
<td>1.00000</td>
<td>1.00000</td>
<td>-1831.</td>
</tr>
</tbody>
</table>

The test following the baseline log-likelihood is the likelihood-ratio equivalent of the regression F-test, testing the significance of all regressors except the constant.

To include the progress of the maximization routine, click on Options when in the Model dialog, and set the Write results for every to one:

When re-estimating the model, iteration output is printed:

Starting values
parameters
0.58338 0.00000

gradients
4.5086e-016 0.017109

Initial function = -0.652349855122

Position after 1 Newton iterations
parameters
-2.7420 0.34407

gradients
0.00081772 0.0080559

function value = -0.649393953514

Position after 2 Newton iterations
parameters
-2.7723 0.34758

gradients
5.8370e-007 5.8958e-006

function value = -0.649392193746

Position after 3 Newton iterations
Status: Strong convergence
parameters
-2.7723 0.34758

gradients
5.1656e-013 5.3271e-012

function value = -0.649392193744
Afterwards reset the Write results for every to zero (or use Reset default on the Model Options dialog).

Next we may wish to inspect the derivatives of the probabilities. Select Test/Further Output, and mark it as follows:

The output also gives a derivative of 0.08, although after rounding the t-value is 4.1 rather than 4.0.

Derivatives of probabilities at sample frequencies
Probabilities:
State 0 0.35816
State 1 0.64184
Derivatives: mean State 0 State 1
Constant 1.0000 0.63730 -0.63730
LINC 9.6649 -0.079902 0.079902
Quasi-elasticities:
State 0 State 1
Constant 0.63730 -0.63730
LINC -0.77225 0.77225
Elasticities:
State 0 State 1
Constant 0.22825 -0.40905
LINC -0.27658 0.49566
t-values:
State 0 State 1
Constant 3.3468 -3.3468
LINC -4.0514 4.0514

### 3.4 Binary probit estimation

A probit model can be estimated by selecting Probit in the model settings. After adding the remaining regressors, the probit and logit results can be compared:

<table>
<thead>
<tr>
<th></th>
<th>logit Coefficient</th>
<th>t-value</th>
<th>probit Coefficient</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-22.6881</td>
<td>-15.2</td>
<td>-12.6035</td>
<td>-15.3</td>
</tr>
<tr>
<td>LINC</td>
<td>2.37962</td>
<td>16.2</td>
<td>1.32838</td>
<td>16.4</td>
</tr>
<tr>
<td>LSIZE</td>
<td>2.76068</td>
<td>19.1</td>
<td>1.56383</td>
<td>19.9</td>
</tr>
<tr>
<td>BUSCAR</td>
<td>-3.03883</td>
<td>-19.5</td>
<td>-1.74894</td>
<td>-20.0</td>
</tr>
<tr>
<td>URBA</td>
<td>-0.117711</td>
<td>-4.20</td>
<td>-0.0660472</td>
<td>-4.09</td>
</tr>
<tr>
<td>HHAGE</td>
<td>-0.126291</td>
<td>-8.24</td>
<td>-0.0780880</td>
<td>-8.80</td>
</tr>
</tbody>
</table>
3.5 Grouped logit estimation

The difference in the coefficients is largely due to the different scaling implicit in the two models. The derivatives of the probabilities show that the two models are actually very close:

<table>
<thead>
<tr>
<th></th>
<th>logit Derivative</th>
<th>t-value</th>
<th>probit Derivative</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>5.2155</td>
<td>15.2</td>
<td>4.5445</td>
<td>15.3</td>
</tr>
<tr>
<td>LINC</td>
<td>−0.54700</td>
<td>−16.2</td>
<td>−0.47898</td>
<td>−16.4</td>
</tr>
<tr>
<td>LSIZE</td>
<td>−0.63463</td>
<td>−19.1</td>
<td>−0.56388</td>
<td>−19.9</td>
</tr>
<tr>
<td>BUSCAR</td>
<td>0.69857</td>
<td>19.5</td>
<td>0.63063</td>
<td>19.8</td>
</tr>
<tr>
<td>URBA</td>
<td>0.027060</td>
<td>4.20</td>
<td>0.023815</td>
<td>4.10</td>
</tr>
<tr>
<td>HHAGE</td>
<td>0.029032</td>
<td>8.24</td>
<td>0.028157</td>
<td>8.77</td>
</tr>
</tbody>
</table>

3.5 Grouped logit estimation

The grouped data are provided with PcGive as NLcargrouped.in7:

Here \( m_i \) is the number of observations with car in each income class, and \( n - m \) the number without car. To estimate the model on grouped data, add these two variables as \( Y \) variable:

Otherwise, the model formulation and estimation procedure is unchanged. The output is:

CS( 3) Modelling \( m - n - m \) by grouped LOGIT
The dataset is: .\OxMetrics7\data\NLcargrouped.in7
The estimation sample is 1 - 5

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.91536</td>
<td>3.48</td>
<td>0.040</td>
</tr>
</tbody>
</table>
Chapter 3 Tutorial on Discrete Choice Modelling

LINC  
-0.361811  0.08673  -4.17  0.025

log-likelihood  
-1830.88379  no. of states  2
no. of observations  5  no. of parameters  2
baseline log-lik  
-1839.627  Test: Chi^2( 1)  17.486 [0.000]**
AIC  3665.76758  AIC/T  733.153516
Newton estimation (eps1=0.0001; eps2=0.005): Strong convergence

<table>
<thead>
<tr>
<th>Count</th>
<th>Frequency</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>State 0</td>
<td>1810</td>
<td>0.64184</td>
</tr>
<tr>
<td>State 1</td>
<td>1010</td>
<td>0.35816</td>
</tr>
<tr>
<td>Total</td>
<td>2820</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

3.6 Multinomial logit estimation

Estimating the multinomial logit model only requires changing the explanatory variable from CAR01 to CAR:

CS( 4) Modelling CAR by LOGIT
The estimation sample is 1 - 2820

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant(S1)</td>
<td>-16.0707</td>
<td>1.657</td>
<td>-9.70</td>
</tr>
<tr>
<td>LINC(S1)</td>
<td>1.68986</td>
<td>0.1636</td>
<td>10.3</td>
</tr>
<tr>
<td>LSIZE(S1)</td>
<td>2.55517</td>
<td>0.1624</td>
<td>15.7</td>
</tr>
<tr>
<td>BUSCAR(S1)</td>
<td>-3.00298</td>
<td>0.1866</td>
<td>-16.1</td>
</tr>
<tr>
<td>URBA(S1)</td>
<td>-0.129213</td>
<td>0.03121</td>
<td>-4.14</td>
</tr>
<tr>
<td>HHAGE(S1)</td>
<td>-0.195673</td>
<td>0.01815</td>
<td>-10.8</td>
</tr>
<tr>
<td>Constant(S2)</td>
<td>-29.5708</td>
<td>1.795</td>
<td>-16.5</td>
</tr>
<tr>
<td>LINC(S2)</td>
<td>2.94892</td>
<td>0.1757</td>
<td>16.8</td>
</tr>
<tr>
<td>LSIZE(S2)</td>
<td>2.68602</td>
<td>0.1707</td>
<td>15.7</td>
</tr>
<tr>
<td>BUSCAR(S2)</td>
<td>-3.02350</td>
<td>0.2100</td>
<td>-14.4</td>
</tr>
<tr>
<td>URBA(S2)</td>
<td>-0.113461</td>
<td>0.03312</td>
<td>-3.43</td>
</tr>
<tr>
<td>HHAGE(S2)</td>
<td>-0.0582466</td>
<td>0.01816</td>
<td>-3.21</td>
</tr>
<tr>
<td>Constant(S3)</td>
<td>-49.4148</td>
<td>3.038</td>
<td>-16.3</td>
</tr>
<tr>
<td>LINC(S3)</td>
<td>4.51794</td>
<td>0.2910</td>
<td>15.5</td>
</tr>
<tr>
<td>LSIZE(S3)</td>
<td>5.84283</td>
<td>0.3278</td>
<td>17.8</td>
</tr>
<tr>
<td>BUSCAR(S3)</td>
<td>-3.59800</td>
<td>0.3693</td>
<td>-9.74</td>
</tr>
<tr>
<td>URBA(S3)</td>
<td>-0.0798822</td>
<td>0.05506</td>
<td>-1.45</td>
</tr>
<tr>
<td>HHAGE(S3)</td>
<td>-0.0444871</td>
<td>0.03690</td>
<td>-1.21</td>
</tr>
</tbody>
</table>

log-likelihood  
-2874.89943  no. of states  4
no. of observations  2820  no. of parameters  18
baseline log-lik  
-3528.374  Test: Chi^2( 15)  1306.9 [0.000]**
AIC  5785.79886  AIC/T  2.05170172
mean(CAR)  1.01099  var(CAR)  0.851298
Newton estimation (eps1=0.0001; eps2=0.005): Strong convergence

<table>
<thead>
<tr>
<th>Count</th>
<th>Frequency</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>State 0</td>
<td>1010</td>
<td>0.35816</td>
</tr>
<tr>
<td>State 1</td>
<td>944</td>
<td>0.33475</td>
</tr>
<tr>
<td>State 2</td>
<td>691</td>
<td>0.24504</td>
</tr>
<tr>
<td>State 3</td>
<td>175</td>
<td>0.06206</td>
</tr>
<tr>
<td>Total</td>
<td>2820</td>
<td>1.00000</td>
</tr>
</tbody>
</table>
The derivatives can again be obtained from Test/Further Output. Additional results can be obtained from Test/Further Output, for probabilities at the sample mean of the regressors. To compute probabilities for other households, use Test/Norm Observation. For example, for a household similar to household A:

resulting in probabilities:

Derivatives of probabilities at specified values
Probabilities:
State 0 0.13869
State 1 0.74990
State 2 0.094391
State 3 0.017023

Derivatives:

<table>
<thead>
<tr>
<th>Variable</th>
<th>State 0</th>
<th>State 1</th>
<th>State 2</th>
<th>State 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.0000</td>
<td>2.1751</td>
<td>-0.29014</td>
<td>-1.3108</td>
</tr>
<tr>
<td>LINC</td>
<td>8.7000</td>
<td>-0.22502</td>
<td>0.050525</td>
<td>0.12520</td>
</tr>
<tr>
<td>LSIZE</td>
<td>1.4000</td>
<td>-0.31470</td>
<td>0.21451</td>
<td>0.039351</td>
</tr>
<tr>
<td>BUSCAR</td>
<td>0.0000</td>
<td>0.36039</td>
<td>-0.30327</td>
<td>-0.040110</td>
</tr>
<tr>
<td>URBA</td>
<td>1.0000</td>
<td>0.015112</td>
<td>-0.015183</td>
<td>-0.00042430</td>
</tr>
<tr>
<td>HHAGE</td>
<td>2.0000</td>
<td>0.021218</td>
<td>-0.032008</td>
<td>0.0089429</td>
</tr>
</tbody>
</table>

Test/Norm Observation can also graph the probabilities at the sample regressor mean (or selected values). For each variable in turn the value is varied over the sample range, while keeping the others fixed at the sample mean (or user selected value). Figure 3.3 illustrates. The first graph shows that at an annual income of Fl.1000 nearly 100% of households has no car (‘NONE’). At the other end of the income spectrum, at Fl.100 000, about 60% has one car or more (‘MORE’), with most of the remainder made up by a new car (‘NEW’). This is while keeping the other regressors at their sample mean.

Finally, Test/Graphic Analysis allows for further graphical inspection of the estimated model:
The first four graph plot the histograms of the estimated probabilities for each state: $\hat{p}_{is}$, $i = 1, \ldots, n$, $s = 1, \ldots, S$. The next four graphs plot the histograms of the estimated probabilities of the observed state: $\hat{p}_{is} | y_i = s$ by state. The final graph plots the histogram of all $\hat{p}_{is} | y_i = s$ together.

### 3.7 Conditional logit estimation

Conditional logit estimation uses the Z variable type. As an example we replicate a normal logit estimation as conditional logit.

First estimate a multinomial regression model with CAR as dependent variable, and a constant and LINC as regressor:

```
CS(5) Modelling CAR by LOGIT
```
3.7 Conditional logit estimation

The estimation sample is 1 - 2820

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant(S1)</td>
<td>1.37083</td>
<td>0.9684</td>
<td>1.42</td>
</tr>
<tr>
<td>LINC(S1)</td>
<td>-0.149804</td>
<td>0.1007</td>
<td>-1.49</td>
</tr>
<tr>
<td>Constant(S2)</td>
<td>-9.92892</td>
<td>1.088</td>
<td>-9.13</td>
</tr>
<tr>
<td>LINC(S2)</td>
<td>0.982743</td>
<td>0.1117</td>
<td>8.80</td>
</tr>
<tr>
<td>Constant(S3)</td>
<td>-8.06614</td>
<td>1.762</td>
<td>-4.58</td>
</tr>
<tr>
<td>LINC(S3)</td>
<td>0.652000</td>
<td>0.1809</td>
<td>3.60</td>
</tr>
</tbody>
</table>

log-likelihood: -3466.80131
no. of states: 4
no. of observations: 2820
no. of parameters: 6
baseline log-lik: -3528.374
Test: Chi^2(3) = 123.15 [0.0000]**

For the next step, create a variable called zero in the database, which just has value zero for all observations. The alternative independent LINC regressor can be implemented as:

\[
\begin{align*}
V_{i0} &= 0, \\
V_{i1} &= \beta_{11} + 0 + \gamma_1 \text{LINC} + 0 + 0, \\
V_{i2} &= \beta_{12} + 0 + 0 + \gamma_2 \text{LINC} + 0, \\
V_{i3} &= \beta_{13} + 0 + 0 + 0 + \gamma_3 \text{LINC}.
\end{align*}
\]

which uses the first state (alternative 0) as reference state. When a constant is added, the resulting specification in terms of (2.3) is:
So each alternative dependent regressor $Z_i$ consists of $S$ variables. In the model formulation stage (this requires Model/Multinomial Discrete Choice), the $Z$ variables must be entered in blocks of size $S$. For the current model this is:

\[
\begin{align*}
Z_1 &= [0, LINC, 0, 0], \\
Z_2 &= [0, 0, 0, LINC],
\end{align*}
\]

which results in the same coefficient estimates:

CS(6) Modelling CAR by LOGIT

The estimation sample is 1 - 2820

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant(S1)</td>
<td>1.37083</td>
<td>0.9684</td>
<td>1.42</td>
</tr>
<tr>
<td>Constant(S2)</td>
<td>-9.92892</td>
<td>1.088</td>
<td>-9.13</td>
</tr>
<tr>
<td>Constant(S3)</td>
<td>-8.06614</td>
<td>1.762</td>
<td>-4.58</td>
</tr>
<tr>
<td>Z1</td>
<td>-0.149804</td>
<td>0.1007</td>
<td>-1.49</td>
</tr>
<tr>
<td>Z2</td>
<td>0.982743</td>
<td>0.1117</td>
<td>8.80</td>
</tr>
<tr>
<td>Z3</td>
<td>0.652000</td>
<td>0.1809</td>
<td>3.60</td>
</tr>
</tbody>
</table>

log-likelihood: -3466.80131  no. of states: 4
no. of observations: 2820  no. of parameters: 6
baseline log-lik: -3528.374  Test: Chi$^2$(3) = 123.15 [0.0000]**

$Z_1 = [\text{zero}, \text{LINC}, \text{zero}, \text{zero}]$

$Z_2 = [\text{zero}, \text{zero}, \text{LINC}, \text{zero}]$

$Z_3 = [\text{zero}, \text{zero}, \text{zero}, \text{LINC}]$

Note that this technique of writing a multinomial logit as a conditional logit model can be used to impose certain type of restrictions. For example, $\gamma_1 = 0$ is imposed by deleting the first $Z$ variable:

\[
Z_1 = [0, 0, \text{LINC}, 0],
\]

and then imposing $\gamma_2 = \gamma_3$ by:

\[
Z = [0, 0, \text{LINC}, \text{LINC}].
\]
3.8 Automatic model selection for binary logit and probit

The process of model simplification can be surprisingly time consuming, so we have developed an automated procedure called Autometrics, see Doornik (2009) and Volume I. The objective is to let the computer do a large part of what was previously done by hand. Autometrics is a computer implementation of general-to-specific modelling. Because it has been developed for a general maximum-likelihood setting, it can be applied to binary logit and probit as well. The main difference compared to regression is the lack of diagnostic tests. Otherwise the principle is very much the same, with likelihood-ratio tests replacing $F$-tests.

As an illustration we return to the example using CAR01. First create the squares of HHAGE, SIZE and INC/1000. We have called these HHAGEsqr, SIZEsqr, INCsqr respectively. This is the Algebra code:

\[
\begin{align*}
\text{HHAGEsqr} &= \text{HHAGE}^2; \\
\text{SIZEsqr} &= \text{SIZE}^2; \\
\text{INCsqr} &= (\text{INC}/1000)^2;
\end{align*}
\]

Now formulate an initial model (the GUM, general unrestricted model) which has all the variables in (except for CAR, of course). This initial model has four insignificant variables, and we’ll let Autometrics do the variable selection at 5%. The procedure finds two possible terminal models:

<table>
<thead>
<tr>
<th>p-values in Final GUM and terminal model(s)</th>
<th>Final GUM</th>
<th>terminal 1</th>
<th>terminal 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINC</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
</tr>
<tr>
<td>SIZE</td>
<td>0.27647689</td>
<td>0.00000000</td>
<td>.</td>
</tr>
<tr>
<td>BUSCAR</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
</tr>
<tr>
<td>URBA</td>
<td>0.00003613</td>
<td>0.00003657</td>
<td>0.00003273</td>
</tr>
<tr>
<td>HHAGE</td>
<td>0.00223200</td>
<td>0.00153731</td>
<td>0.00408628</td>
</tr>
<tr>
<td>LSIZE</td>
<td>0.0000025</td>
<td>0.00000000</td>
<td>0.00000000</td>
</tr>
<tr>
<td>SIZEsqr</td>
<td>0.87094782</td>
<td>.</td>
<td>0.00000000</td>
</tr>
<tr>
<td>HHAGEsqr</td>
<td>0.00000983</td>
<td>0.00000480</td>
<td>0.00001805</td>
</tr>
<tr>
<td>$k$</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>parameters</td>
<td>9</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>loglik</td>
<td>-1333.5</td>
<td>-1333.5</td>
<td>-1334.1</td>
</tr>
<tr>
<td>SC</td>
<td>0.97109</td>
<td>0.96828</td>
<td>0.96872</td>
</tr>
</tbody>
</table>

Terminal 1 is selected on the Schwarz criterion, but there is really not much between the two terminals. The final model is reported below, indicating non-linear effects from household size and the age of the head of household:

Summary of Autometrics search

<table>
<thead>
<tr>
<th>initial search space</th>
<th>2(^{10})</th>
<th>final search space</th>
<th>2(^{8})</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. estimated models</td>
<td>17</td>
<td>no. terminal models</td>
<td>2</td>
</tr>
<tr>
<td>test form</td>
<td>LR-Chi(^{2})</td>
<td>target size</td>
<td>Default:0.05</td>
</tr>
<tr>
<td>outlier detection</td>
<td>no</td>
<td>presearch reduction</td>
<td>lags</td>
</tr>
<tr>
<td>backtesting</td>
<td>GUM0</td>
<td>tie-breaker</td>
<td>SC</td>
</tr>
<tr>
<td>diagnostics p-value</td>
<td>0.01</td>
<td>search effort</td>
<td>standard</td>
</tr>
<tr>
<td>time</td>
<td>0.56</td>
<td>Autometrics version</td>
<td>1.5e</td>
</tr>
</tbody>
</table>
CS(26) Modelling CAR01 by PROBIT

The dataset is: D:\OxMetrics7\data\NLcar.in7
The estimation sample is 1 - 2820

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-12.2280</td>
<td>0.8328</td>
<td>-14.7</td>
</tr>
<tr>
<td>LINC</td>
<td>1.26826</td>
<td>0.08247</td>
<td>15.4</td>
</tr>
<tr>
<td>SIZE</td>
<td>-0.510694</td>
<td>0.08459</td>
<td>-6.04</td>
</tr>
<tr>
<td>BUSCAR</td>
<td>-1.74640</td>
<td>0.08703</td>
<td>-20.1</td>
</tr>
<tr>
<td>URBA</td>
<td>-0.0671979</td>
<td>0.01625</td>
<td>-4.13</td>
</tr>
<tr>
<td>HHAGE</td>
<td>0.164265</td>
<td>0.05181</td>
<td>3.17</td>
</tr>
<tr>
<td>LSIZE</td>
<td>2.48777</td>
<td>0.1962</td>
<td>12.7</td>
</tr>
<tr>
<td>HHAGEsqr</td>
<td>-0.0164389</td>
<td>0.003587</td>
<td>-4.58</td>
</tr>
</tbody>
</table>

log-likelihood -1333.50092 no. of states 2
no. of observations 2820 no. of parameters 8
zeroline log-lik -1954.675 Test: Chi^2( 8) 1242.3 [0.0000]**
AIC 2683.00184 AIC/n 0.951419093
mean(CAR01) 0.641844 var(CAR01) 0.22988
Newton estimation (eps1=0.0001; eps2=0.005): Strong convergence

<table>
<thead>
<tr>
<th>Count</th>
<th>Frequency</th>
<th>Probability</th>
<th>loglik</th>
</tr>
</thead>
<tbody>
<tr>
<td>State 0</td>
<td>1010</td>
<td>0.35816</td>
<td>0.35728</td>
</tr>
<tr>
<td>State 1</td>
<td>1810</td>
<td>0.64184</td>
<td>0.64272</td>
</tr>
<tr>
<td>Total</td>
<td>2820</td>
<td>1.00000</td>
<td>1.00000</td>
</tr>
</tbody>
</table>
Part III

Panel Data Models (DPD)

with Manuel Arellano and Stephen Bond
4.1 Introduction

In a panel data setting, we have time-series observations on multiple entities, for example companies or individuals. We denote the cross-section sample size by $N$, and, in an ideal setting, have $t = 1, \ldots, T$ time-series observations covering the same calendar period. This is called a balanced panel. In practice, it often happens that some cross-sections start earlier, or finish later. The panel data procedures in PcGive are expressly designed to handle such unbalanced panels.

When $T$ is large, $N$ small, and the panel balanced, it will be possible to use the simultaneous-equations modelling facilities of PcGive. When $T$ is small, and the model dynamic (i.e. includes a lagged dependent variable), the estimation bias can be substantial (see Nickell, 1981). Methods to cope with such dynamic panel data models are the primary focus of this part of PcGive, particularly the GMM-type estimators of Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998), but also some of the Anderson and Hsiao (1982) estimators. Additional information can be found in Arellano (2003). In addition, PcGive makes several of the standard static panel data estimators available, such as between and within groups and feasible GLS. Table 4.1 lists the available estimators.

<table>
<thead>
<tr>
<th>static panel data models</th>
<th>dynamic panel data models</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS in levels</td>
<td>OLS in levels</td>
</tr>
<tr>
<td>Between estimator</td>
<td>one-step IV estimation (Anderson–Hsiao IV)</td>
</tr>
<tr>
<td>Within estimator</td>
<td>one-step GMM estimation</td>
</tr>
<tr>
<td>Feasible GLS</td>
<td>one-step estimation with robust std. errors</td>
</tr>
<tr>
<td>GLS (OLS residuals)</td>
<td>two-step estimation</td>
</tr>
<tr>
<td>Maximum likelihood (ML)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1  Static and dynamic panel data estimators of PcGive
This chapter only provides a cursory overview of panel data methods. In addition to the referenced literature, there are several text books on panel data, notably Hsiao (1986) and more recently Baltagi (1995). Especially relevant is the book by Arellano (2003). Many general econometric text books have a chapter on panel data (but rarely treating dynamic models), see, e.g., Johnston and DiNardo (1997, Ch. 12).

4.2 Econometric methods for static panel data models

4.2.1 Static panel-data estimation

The static single-equation panel model can be written as:

$$y_{it} = x_{it}' \gamma + \lambda_t + \eta_i + v_{it}, \quad t = 1, \ldots, T, \quad i = 1, \ldots, N. \quad (4.1)$$

The $\lambda_t$ and $\eta_i$ are respectively time and individual specific effects and $x_{it}$ is a $k^*$ vector of explanatory variables. $N$ is the number of cross-section observations. The total number of observations is $NT$.

Stacking the data for an individual according to time yields:

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} x_{i1}' \gamma \\ x_{i2}' \gamma \\ \vdots \\ x_{iT}' \gamma \end{bmatrix} + \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_T \end{bmatrix} + \begin{bmatrix} \eta_i \\ \eta_i \\ \vdots \\ \eta_i \end{bmatrix} + \begin{bmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{iT} \end{bmatrix},$$

or, more concisely:

$$y_i = X_i \gamma + \lambda + \eta_i + v_i, \quad i = 1, \ldots, N,$$

where $y_i$, $X_i$, and $\lambda_i$ are $T \times 1$, and $\eta_i$ is a $T$ column of ones. Using $D_i$ for the matrix with dummies:

$$y_i = X_i \gamma + D_i \delta + v_i, \quad i = 1, \ldots, N.$$

The next step is to stack all individuals, combining the data into $W = [X : D]$:

$$y = W\beta + v. \quad (4.2)$$

Baltagi (1995) reviews the standard estimation methods used in this setting:

- **OLS** estimation.
- **LSDV** (least squares dummy variables) estimation uses individual dummies in the OLS regression.
- **Within** estimation replaces $y$ and $W$ by deviations from time means (i.e., subtracting the means of each time series).
- **Between** estimation replaces $y$ and $W$ by the means of each individual (leaving $N$ observations).
- **Feasible GLS** (generalized least squares) estimation replaces $y$ and $W$ by deviations from weighted time means. The outcome depends on the choice of the weight, $\theta$.
- **ML** (maximum likelihood) estimation obtains $\theta$ by iterating the GLS procedure. More detail is provided in §7.3.
4.3 Econometric methods for dynamic panel data models

The general model that can be estimated with PcGive is a single equation with individual effects of the form:

\[ y_{it} = \sum_{k=1}^{p} \alpha_k y_{i,t-k} + \beta'(L)x_{it} + \lambda_t + \eta_i + \nu_{it}, \quad t = q + 1, \ldots, T_i; \quad i = 1, \ldots, N, \quad (4.3) \]

where \( \eta_i \) and \( \lambda_t \) are respectively individual and time specific effects, \( x_{it} \) is a vector of explanatory variables, \( \beta(L) \) is a vector of associated polynomials in the lag operator and \( q \) is the maximum lag length in the model. The number of time periods available on the \( i \)th individual, \( T_i \), is small and the number of individuals, \( N \), is large. Identification of the model requires restrictions on the serial correlation properties of the error term \( \nu_{it} \) and/or on the properties of the explanatory variables \( x_{it} \). It is assumed that if the error term was originally autoregressive, the model has been transformed so that the coefficients \( \alpha \)'s and \( \beta \)'s satisfy some set of common factor restrictions. Thus only serially uncorrelated or moving-average errors are explicitly allowed. The \( x_{it} \) are assumed to be independently distributed across individuals with zero mean, but arbitrary forms of heteroscedasticity across units and time are possible. The \( x_{it} \) may or may not be correlated with the individual effects \( \eta_i \), and for each of these cases they may be strictly exogenous, predetermined or endogenous variables with respect to \( \nu_{it} \). A case of particular interest is where the levels \( x_{it} \) are correlated with \( \eta_i \) but where \( \Delta x_{it} \) (and possibly \( \Delta y_{it} \)) are uncorrelated with \( \eta_i \); this allows the use of (suitably lagged) \( \Delta x_{is} \) (and possibly \( \Delta y_{is} \)) as instruments for equations in levels.

The \( (T_i - q) \) equations for individual \( i \) can be written conveniently in the form:

\[ y_i = W_i \delta + \iota_i \eta_i + v_i, \]

where \( \delta \) is a parameter vector including the \( \alpha_k \)'s, the \( \beta \)'s and the \( \lambda \)'s, and \( W_i \) is a data matrix containing the time series of the lagged dependent variables, the \( x \)'s and the time dummies. Lastly, \( \iota_i \) is a \( (T_i - q) \times 1 \) vector of ones. PcGive can be used to compute various linear GMM estimators of \( \delta \) with the general form:

\[ \hat{\delta} = \left( \sum_i W_i^*Z_i \right) A_N \left( \sum_i Z_i^tW_i^* \right)^{-1} \left( \sum_i W_i^*Z_i \right) A_N \left( \sum_i Z_i^t\gamma_i^* \right), \]

where

\[ A_N = \left( \frac{1}{N} \sum_i Z_i^tH_iZ_i \right)^{-1}, \]

and \( W_i^* \) and \( \gamma_i^* \) denote some transformation of \( W_i \) and \( y_i \) (e.g. levels, first differences, orthogonal deviations, combinations of first differences (or orthogonal deviations) and levels, deviations from individual means). \( Z_i \) is a matrix of instrumental variables.
which may or may not be entirely internal, and $H_i$ is a possibly individual-specific weighting matrix.

If the number of columns of $Z_i$ equals that of $W_i^*$, $A_N$ becomes irrelevant and $\hat{\delta}$ reduces to

$$\hat{\delta} = \left( \sum_i Z_i' W_i^* \right)^{-1} \left( \sum_i Z_i' y_i^* \right).$$

In particular, if $Z_i = W_i^*$ and the transformed $W_i$ and $y_i$ are deviations from individual means or orthogonal deviations, then $\hat{\delta}$ is the within-groups estimator. As another example, if the transformation denotes first differences, $Z_i = I_{T_i} \otimes x_i'$ and $H_i = \hat{v}_i^* \hat{v}_i'^*$, where the $\hat{v}_i^*$ are some consistent estimates of the first-differenced residuals, then $\hat{\delta}$ is the generalized three-stage least squares estimator of Chamberlain (1984). These two estimators require the $x_{it}$ to be strictly exogenous with respect to $v_{it}$ for consistency. In addition, the within-groups estimator can only be consistent as $N \to \infty$ for fixed $T$ if $W_i^*$ does not contain lagged dependent variables and all the explanatory variables are strictly exogenous.

When estimating dynamic models, we shall therefore typically be concerned with transformations that allow the use of lagged endogenous (and predetermined) variables as instruments in the transformed equations. Efficient GMM estimators will typically exploit a different number of instruments in each time period. Estimators of this type are discussed in Arellano (1988), Arellano and Bond (1991), Arellano and Bover (1995) and Blundell and Bond (1998). PcGive can be used to compute a range of linear GMM estimators of this type.

Where there are no instruments available that are uncorrelated with the individual effects $\eta_i$, the transformation must eliminate this component of the error term. The first difference and orthogonal deviations transformations are two examples of transformations that eliminate $\eta_i$ from the transformed error term, without at the same time introducing all lagged values of the disturbances $v_{it}$ into the transformed error term. Hence these transformations allow the use of suitably lagged endogenous (and predetermined) variables as instruments. For example, if the panel is balanced, $p = 1$, there are no explanatory variables nor time effects, the $v_{it}$ are serially uncorrelated, and the initial conditions $y_{i1}$ are uncorrelated with $v_{it}$ for $t = 2, \ldots, T$, then using first differences we have:

1 Orthogonal deviations, as proposed by Arellano (1988) and Arellano and Bover (1995), express each observation as the deviation from the average of future observations in the sample for the same individual, and weight each deviation to standardize the variance, i.e.

$$x_{it}^* = \left( x_{it} - \frac{x_{i(t+1)} + \ldots + x_{iT}}{T-t} \right) \left( \frac{T-t}{T-t+1} \right)^{1/2}$$

for $t = 1, \ldots, T - 1$.

If the original errors are serially uncorrelated and homoscedastic, the transformed errors will also be serially uncorrelated and homoscedastic.

2 There are many other transformations which share these properties. See Arellano and Bover (1995) for further discussion.
### 4.3 Econometric methods for dynamic panel data models

#### Equations and Instruments available

<table>
<thead>
<tr>
<th>Equation</th>
<th>Instruments available</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta y_{i3} = \alpha \Delta y_{i2} + \Delta v_{i3}$</td>
<td>$y_{i1}$</td>
</tr>
<tr>
<td>$\Delta y_{i4} = \alpha \Delta y_{i3} + \Delta v_{i4}$</td>
<td>$y_{i1}, y_{i2}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$\Delta y_{iT} = \alpha \Delta y_{i,T-1} + \Delta v_{iT}$</td>
<td>$y_{i1}, y_{i2}, \ldots, y_{i,T-2}$</td>
</tr>
</tbody>
</table>

In this case, $y_i^* = (\Delta y_{i3}, \ldots, \Delta y_{iT})'$, $W_i^* = (\Delta y_{i2}, \ldots, \Delta y_{i,T-1})'$ and

$$Z_i = Z_i^D = \begin{pmatrix} y_{i1} & 0 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & y_{i1} & y_{i2} & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & y_{i1} & y_{i2} & \ldots & y_{i,T-2} \end{pmatrix}$$

Notice that precisely the same instrument set would be used to estimate the model in orthogonal deviations. Where the panel is unbalanced, for individuals with incomplete data the rows of $Z_i$ corresponding to the missing equations are deleted, and missing values in the remaining rows are replaced by zeros.

In PcGive, we call one-step estimates those which use some known matrix as the choice for $H_i$. For a first-difference procedure, the one-step estimator uses

$$H_i = H_i^D = \frac{1}{2} \begin{pmatrix} 2 & -1 & \ldots & 0 \\ -1 & 2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & -1 \end{pmatrix},$$

while for a levels or orthogonal deviations procedure, the one-step estimator sets $H_i$ to an identity matrix. If the $v_{it}$ are heteroscedastic, a two-step estimator which uses

$$H_i = \tilde{\hat{v}}_i^\top \tilde{\hat{v}}_i^*,$$

where $\tilde{\hat{v}}_i^*$ are one-step residuals, is more efficient (cf. White, 1982). PcGive produces both one-step and two-step GMM estimators, with asymptotic variance matrices that are heteroscedasticity-consistent in both cases. Users should note that, particularly when the $v_{it}$ are heteroscedastic, simulations suggest that the asymptotic standard errors for the two-step estimators can be a poor guide for hypothesis testing in typical sample sizes. In these cases, inference based on asymptotic standard errors for the one-step estimators seems to be more reliable (see Arellano and Bond, 1991, and Blundell and Bond, 1998 for further discussion).

In models with explanatory variables, $Z_i$ may consist of sub-matrices with the block-diagonal form illustrated above (exploiting all or part of the moment restrictions available), concatenated to straightforward one-column instruments. A judicious choice of the $Z_i$ matrix should strike a compromise between prior knowledge (from economic theory and previous empirical work), the characteristics of the sample and computer
limitations (see Arellano and Bond, 1991 for an extended discussion and illustration). For example, if a predetermined regressor $x_{it}$ correlated with the individual effect, is added to the model discussed above, i.e.

$$E(x_{it}v_{is}) = 0 \text{ for } s \geq t$$

$$\neq 0 \text{ otherwise}$$

$$E(x_{it}\eta_i) \neq 0$$

then the corresponding optimal $Z_i$ matrix is given by

$$Z_i = \begin{pmatrix}
    y_{i1} & x_{i1} & x_{i2} & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\
    0 & 0 & 0 & y_{i1} & y_{i2} & x_{i1} & x_{i2} & x_{i3} & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\
    \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
    0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & y_{i1} & \cdots & y_{i,T-2} & x_{i1} & \cdots & x_{i,T-1}
\end{pmatrix}$$

Where the number of columns in $Z_i$ is very large, computational considerations may require those columns containing the least informative instruments to be deleted. Even when computing speed is not an issue, it may be advisable not to use the whole history of the series as instruments in the later cross-sections. For a given cross-sectional sample size ($N$), the use of too many instruments may result in (small sample) overfitting biases. When overfitting results from the number of time periods ($T$) becoming large relative to the number of individuals ($N$), and there are no endogenous regressors present, these GMM estimators are biased towards within groups, which is not a serious concern since the within-groups estimator is itself consistent for models with predetermined variables as $T$ becomes large (see Álvarez and Arellano, 1998). However, in models with endogenous regressors, using too many instruments in the later cross-sections could result in seriously biased estimates. This possibility can be investigated in practice by comparing the GMM and within-groups estimates.

The assumption of no serial correlation in the $v_{it}$ is essential for the consistency of estimators such as those considered in the previous examples, which instrument the lagged dependent variable with further lags of the same variable. Thus, PcGive reports tests for the absence of first-order and second-order serial correlation in the first-differenced residuals. If the disturbances $v_{it}$ are not serially correlated, there should be evidence of significant negative first-order serial correlation in differenced residuals (i.e. $\hat{v}_{it} - \hat{v}_{i,t-1}$), and no evidence of second-order serial correlation in the differenced residuals. These tests are based on the standardized average residual autocovariances which are asymptotically $\mathcal{N}(0,1)$ variables under the null of no autocorrelation. The tests reported are based on estimates of the residuals in first differences, even when the estimator is obtained using orthogonal deviations. More generally, the Sargan (1964) tests of overidentifying restrictions are also reported. That is, if $A_N$ has been chosen

---

3Although the validity of orthogonality conditions is not affected, the transformation to orthogonal deviations can induce serial correlation in the transformed error term if the $v_{it}$ are serially uncorrelated but heteroscedastic.
optimally for any given $\mathbf{Z}_i$, the statistic

$$
S = \left( \sum_i \mathbf{\hat{v}}'_i \mathbf{Z}_i \right) \mathbf{A}_N \left( \sum_i \mathbf{Z}_i' \mathbf{\hat{v}}_i^* \right)
$$

is asymptotically distributed as a $\chi^2$ with as many degrees of freedom as overidentifying restrictions, under the null hypothesis of the validity of the instruments. Note that only the Sargan test based on the two-step GMM estimator is heteroscedasticity-consistent.

Again, Arellano and Bond (1991) provide a complete discussion of these procedures.

Where there are instruments available that are uncorrelated with the individual effects $\eta_i$, these variables can be used as instruments for the equations in levels. Typically this will imply a set of moment conditions relating to the equations in first differences (or orthogonal deviations) and a set of moment conditions relating to the equations in levels, which need to be combined to obtain the efficient GMM estimator.\footnote{In special cases it may be efficient to use only the equations in levels; for example, in a model with no lagged dependent variables and all regressors strictly exogenous and uncorrelated with individual effects.}

For example, if the simple AR(1) model considered earlier is mean-stationary, then the first differences $\Delta y_{it}$ will be uncorrelated with $\eta_i$, and this implies that $\Delta y_{i,t-1}$ can be used as instruments in the levels equations (see Arellano and Bover, 1995 and Blundell and Bond, 1998 for further discussion). In addition to the instruments available for the first-differenced equations that were described earlier, we then have:

<table>
<thead>
<tr>
<th>Equations</th>
<th>Instruments available</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_{i3} = \alpha y_{i2} + \eta_i + v_{i3}$</td>
<td>$\Delta y_{i2}$</td>
</tr>
<tr>
<td>$y_{i4} = \alpha y_{i3} + \eta_i + v_{i4}$</td>
<td>$\Delta y_{i3}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$y_{iT} = \alpha y_{i,T-1} + \eta_i + v_{iT}$</td>
<td>$\Delta y_{i,T-1}$</td>
</tr>
</tbody>
</table>

Notice that no instruments are available in this case for the first levels equation (i.e., $y_{i2} = \alpha y_{i1} + \eta_i + v_{i2}$), and that using further lags of $\Delta y_{is}$ as instruments here would be redundant, given the instruments that are being used for the equations in first differences. In a balanced panel, we could use only the last levels equation (i.e., $y_{iT} = \alpha y_{i,T-1} + \eta_i + v_{iT}$), where $(\Delta y_{i2}, \Delta y_{i3}, \ldots, \Delta y_{i,T-1})$ would all be valid instruments; however this approach does not extend conveniently to unbalanced panels.

In this case, we use

$$
y_{i}^* = (\Delta y_{i3}, \ldots, \Delta y_{iT}, y_{i3}, \ldots, y_{iT})',
$$

$$
W_{i}^* = (\Delta y_{i2}, \ldots, \Delta y_{i,T-1}, y_{i2}, \ldots, y_{i,T-1})',
$$

and

$$
\mathbf{Z}_i = 
\begin{pmatrix}
\mathbf{D}_i & 0 & \cdots & 0 \\
0 & \Delta y_{i2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \Delta y_{i,T-1}
\end{pmatrix}
$$
where $Z_D^i$ is the matrix of instruments for the equations in first differences, as described above. Again $Z_i$ would be precisely the same if the transformed equations in $y^*_i$ and $W^*_i$ were in orthogonal deviations rather than first differences. In models with explanatory variables, it may be that the levels of some variables are uncorrelated with $\eta_i$, in which case suitably lagged levels of these variables can be used as instruments in the levels equations, and in this case there may be instruments available for the first levels equation.

For the system of equations in first differences and levels, the one-step estimator computed in PcGive uses the weighting matrix

$$H_i = \begin{pmatrix} H_D^i & 0 \\ 0 & \frac{1}{2}I_i \end{pmatrix}$$

where $H_D^i$ is the weighting matrix described above for the first differenced estimator, and $I_i$ is an identity matrix with dimension equal to the number of levels equations observed for individual $i$. For the system of equations in orthogonal deviations and levels, the one-step estimator computed in PcGive sets $H_i$ to an identity matrix with dimension equal to the total number of equations in the system for individual $i$. In both cases the corresponding two-step estimator uses $H_i = \hat{v}_i \hat{v}_i'$. We adopt these particular one-step weighting matrices because they are equivalent in the following sense: for a balanced panel where all the available linear moment restrictions are exploited (i.e., no columns of $Z_i$ are omitted for computational or small-sample reasons), the associated one-step GMM estimators are numerically identical, regardless of whether the first difference or orthogonal deviations transformation is used to construct the system. Notice though that the one-step estimator is asymptotically inefficient relative to the two-step estimator for both of these systems, even if the $v_{it}$ are homoscedastic.\(^5\) Again simulations have suggested that asymptotic inference based on the one-step versions may be more reliable than asymptotic inference based on the two-step versions, even in moderately large samples (see Blundell and Bond, 1998).

The validity of these extra instruments in the levels equations can be tested using the Sargan statistic provided by PcGive. Since the set of instruments used for the equations in first differences (or orthogonal deviations) is a strict subset of that used in the system of first-differenced (or orthogonal deviations) and levels equations, a more specific test of these additional instruments is a Difference Sargan test which compares the Sargan statistic for the system estimator and the Sargan statistic for the corresponding first-differenced (or orthogonal deviations) estimator. Another possibility is to compare these estimates using a Hausman (1978) specification test, which can be computed here by including another set of regressors that take the value zero in the equations in first differences (or orthogonal deviations), and reproduce the levels of the right-hand side

\(^5\)With levels equations included in the system, the optimal weight matrix depends on unknown parameters (for example, the ratio of $\text{var}(\eta_i)$ to $\text{var}(v_{it})$) even in the homoscedastic case.
variables for the equations in levels. The test statistic is then a Wald test of the hypothesis that the coefficients on these additional regressors are jointly zero. Full details of these test procedures can be found in Arellano and Bond (1991) and Arellano (1995).

\[ \begin{pmatrix} 0 & \cdots & 0 & y_{i2} & \cdots & y_{iT-1} \\ \Delta y_{i2} & \cdots & \Delta y_{iT-1} & y_{i2} & \cdots & y_{iT-1} \end{pmatrix} \]
Chapter 5

Tutorial on Static Panel Data Modelling

5.1 Introduction

The example for static panel modelling is based on Grunfeld (1958), which is also used in Baltagi (1995, Ch. 2). The estimated model is an investment equation with \( N = 10 \) and \( T = 20 \):

\[
I_{it} = \beta_0 + \beta_1 C_{it-1} + \beta_2 F_{it-1} + u_{it}, \quad t = 1935, \ldots, 1954, \ i = 1, \ldots, 10 \quad (5.1)
\]

where

- \( I_{it} \) is current gross investment of firm \( i \),
- \( C_{it-1} \) is beginning-of-year capital stock of firm \( i \),
- \( F_{it-1} \) is value of outstanding shares at beginning of year.

Writing \( u_{it} = \eta_i + v_{it} \) corresponds to (4.1):

\[
y_{it} = \beta_0 + \beta' x_{i,t} + \eta_i + v_{it}, \quad t = 1, \ldots, T, \ i = 1, \ldots, N \quad (5.2)
\]

where \( \eta_i \) is a firm-specific effect, and \( v_{it} \) is IID across individuals (not serially correlated, but it may be heteroscedastic), \( x_{i,t} \) may or may not be correlated with \( \eta_i \).

5.2 Data organization

The data set is provided as grunfeld.xls, which is in the PcGive folder. Load this in OxMetrics. The database is ordered by firm, and within firm by year:
5.2 Data organization

<table>
<thead>
<tr>
<th>Year</th>
<th>I</th>
<th>F.I</th>
<th>C.I</th>
<th>Firm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1935</td>
<td>317.6</td>
<td>3078.5</td>
<td>2.8</td>
<td>1</td>
</tr>
<tr>
<td>1936</td>
<td>391.8</td>
<td>4661.7</td>
<td>52.6</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>1953</td>
<td>1304.4</td>
<td>6241.7</td>
<td>1777.3</td>
<td>1</td>
</tr>
<tr>
<td>1954</td>
<td>1486.7</td>
<td>5593.6</td>
<td>2226.3</td>
<td>1</td>
</tr>
<tr>
<td>1935</td>
<td>209.9</td>
<td>1362.4</td>
<td>53.8</td>
<td>2</td>
</tr>
<tr>
<td>1936</td>
<td>355.3</td>
<td>1807.1</td>
<td>50.5</td>
<td>2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>1953</td>
<td>641.0</td>
<td>2031.3</td>
<td>623.6</td>
<td>2</td>
</tr>
<tr>
<td>1954</td>
<td>459.3</td>
<td>2115.5</td>
<td>669.7</td>
<td>2</td>
</tr>
<tr>
<td>etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OxMetrics does not have special facilities for panel data management, and this is the general format for all data sets. Although it is not necessary, it is useful to have an additional variable which indicates the firm or individual by a number, see §6.2. Figure 5.1 shows the data.

![Figure 5.1 Grunfeld(1958) investment data](image)

Here we have added the Firm variable using the insample command:

```plaintext
firm = insample( 1,1, 20,1) ? 1;
firm = insample( 21,1, 40,1) ? 2;
firm = insample( 41,1, 60,1) ? 3;
firm = insample( 61,1, 80,1) ? 4;
firm = insample( 81,1,100,1) ? 5;
firm = insample(101,1,120,1) ? 6;
firm = insample(121,1,140,1) ? 7;
firm = insample(141,1,160,1) ? 8;
firm = insample(161,1,180,1) ? 9;
firm = insample(181,1,200,1) ? 10;
```
OLS estimation of (5.1) can be done using the Single-equation Dynamic Modelling using PcGive:

EQ(1) Modelling I by OLS (using grunfeld.xls)
The estimation sample is: 1 - 200

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
<th>Part.R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-42.7144</td>
<td>9.512</td>
<td>-4.49</td>
<td>0.000</td>
</tr>
<tr>
<td>F_1</td>
<td>0.115562</td>
<td>0.005836</td>
<td>19.8</td>
<td>0.000</td>
</tr>
<tr>
<td>C_1</td>
<td>0.230678</td>
<td>0.02548</td>
<td>9.05</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[
\text{sigma} = 94.4084 \quad \text{RSS} = 1755850.48 \\
R^2 = 0.812408 \quad F(2,197) = 426.6 \quad [0.000]** \\
\text{log-likelihood} = -908.015 \quad \text{DW} = 0.358 \\
\text{no. of observations} = 200 \quad \text{no. of parameters} = 3 \\
\text{mean(I)} = 145.958 \quad \text{var(I)} = 46799.7
\]

5.3 Static panel data estimation

We now assume that you have loaded the grunfeld.xls data set into OxMetrics. Start modelling in OxMetrics, selecting Models for panel data and model class Static Panel Methods:

Next, select Formulate and formulate the model with I as dependent variable and \( F_1 \) and \( C_1 \) as regressors. The intercept will be added in the next step. PcGive must know about the panel structure of the data. In this case, the easiest way is to define the Year variable: add Year to the model, and mark it as such by right-clicking on it in the model, and selecting R: Year:
Click on OK, to see the Model Settings dialog. Switch off robust standard errors for the duration of this chapter:

The default adds a constant term, so keep that, and press OK again:
Starting with pooled regression replicates the earlier results:

DPD( 1) Modelling I by OLS (using grunfeld.xls)

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>F_1</td>
<td>0.115562</td>
<td>0.005836</td>
<td>19.8</td>
<td>0.000</td>
</tr>
<tr>
<td>C_1</td>
<td>0.230678</td>
<td>0.02548</td>
<td>9.05</td>
<td>0.000</td>
</tr>
<tr>
<td>Constant</td>
<td>-42.7144</td>
<td>9.512</td>
<td>-4.49</td>
<td>0.000</td>
</tr>
</tbody>
</table>

sigma | 94.4084 | sigma^2 | 8912.947 |

R^2 | 0.812408 |

RSS | 1755850.4841 | TSS | 9359943.9289 |

no. of observations | 200 | no. of parameters | 3 |

Warning: standard errors not robust to heteroscedasticity

Transformation used: none

constant: yes | time dummies: | 0 |

number of individuals | 10 (derived from year) |
longest time series | 20 [1935 - 1954] |
shortest time series | 20 (balanced panel) |

Wald (joint): Chi^2(2) = 853.2 [0.000] **
Wald (dummy): Chi^2(1) = 20.17 [0.000] **
AR(1) test: N(0,1) = 28.49 [0.000] **
AR(2) test: N(0,1) = 23.47 [0.000] **

The estimated coefficients and equation standard error (listed as sigma) are identical. Based on the Year variable, PcGive has also listed some information on the panel, namely \( N = 10 \), \( T = 20 \) and that the panel is balanced. The first Wald test is for the significance on all variables except the dummy (which is the constant term), so is the \( \chi^2 \) equivalent to the overall F-test. The next Wald test reports the significance of the constant term, and is just the squared t-value here. The two remaining tests are for first and second order serial correlation, which are both significant when one variable is involved.

Figure 5.2 shows actual and fitted values and residuals for several estimates. The first two graphs are for OLS on the pooled data, which ignores the panel aspect of the data: the firm-specific effect is not at all picked up by the model. Taking first differences of (5.2) removes \( \eta_i \), as well as the intercept \( \beta_0 \). The graphical analysis is in graphs 5.2c and d. The least-squares dummy variable and within-groups methods lead to identical results, except that with LSDV the coefficients on the firm dummies are reported, while in within-groups estimation the regression is performed in deviation from the means for each firm. The final two graphs are for between-groups estimation. This uses the means for each firm, hence has only ten observations.

Note, however, that if a constant or individual dummies are added after taking first differences, these will not have coefficient zero, but instead will estimate trend effects. This is the case in the output reported below, where the constant is not zero (although insignificant). When estimating by OLS on differences, the dummies are added after the transformation, although this default can be changed in the Model Settings dialog. See §7.1 on how the dummies are treated in general.
We finish with Table 5.1 which produces the coefficient estimates for each choice of estimator. It closely reproduces Table 2.1 in Baltagi (1995).

There is some redundancy in the available estimators: when individual dummies are added to the model, OLS (pooled), LSDV and within-groups estimation will all give the same coefficient estimates. The $R^2$ is different for within-groups, because it is based on the dependent variable after removing the individual means.
Table 5.1 Static panel estimates for Grunfeld data

<table>
<thead>
<tr>
<th>Method</th>
<th>$\beta_1$ (SE)</th>
<th>$\beta_2$ (SE)</th>
<th>$\beta_0$ (t-stat)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OLS (pooled regression)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>0.115562 (0.005836)</td>
<td>0.230678 (0.02548)</td>
<td>−42.7144 (9.512)</td>
</tr>
<tr>
<td><strong>OLS on differences</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS-diff</td>
<td>0.0897625 (0.008364)</td>
<td>0.291767 (0.05375)</td>
<td>−1.81889 (3.566)</td>
</tr>
<tr>
<td><strong>LSDV (fixed effects)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSDV</td>
<td>0.110124 (0.01186)</td>
<td>0.310065 (0.01735)</td>
<td>−70.2967 (49.71)</td>
</tr>
<tr>
<td><strong>Within-groups estimation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>0.110124 (0.01186)</td>
<td>0.310065 (0.01735)</td>
<td></td>
</tr>
<tr>
<td><strong>Between-groups estimation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between</td>
<td>0.134646 (0.02875)</td>
<td>0.0320315 (0.1909)</td>
<td>−8.52711 (47.52)</td>
</tr>
<tr>
<td><strong>GLS (using within/between)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLS(w/b)</td>
<td>0.109781 (0.01049)</td>
<td>0.308113 (0.01718)</td>
<td>−57.8344 (28.90)</td>
</tr>
<tr>
<td><strong>GLS (using OLS residuals)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLS</td>
<td>0.109734 (0.01029)</td>
<td>0.307647 (0.01724)</td>
<td>−57.6546 (26.46)</td>
</tr>
<tr>
<td><strong>Maximum likelihood estimation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td>0.109763 (0.01042)</td>
<td>0.307942 (0.01720)</td>
<td>−57.7672 (27.91)</td>
</tr>
</tbody>
</table>
Chapter 6

Tutorial on Dynamic Panel Data Modelling

6.1 Introduction

The applications in this chapter use the data from Arellano and Bond (1991), consisting of an unbalanced panel of 140 UK companies. The variables in the database are:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IND</td>
<td>industry indicator</td>
</tr>
<tr>
<td>YEAR</td>
<td>year</td>
</tr>
<tr>
<td>EMP</td>
<td>employment</td>
</tr>
<tr>
<td>WAGE</td>
<td>real product wage</td>
</tr>
<tr>
<td>CAP</td>
<td>gross capital</td>
</tr>
<tr>
<td>INDOUPT</td>
<td>industry output</td>
</tr>
<tr>
<td>log(EMP)</td>
<td></td>
</tr>
<tr>
<td>log(WAGE)</td>
<td></td>
</tr>
<tr>
<td>log(CAP)</td>
<td></td>
</tr>
<tr>
<td>log(INDOUTPT)</td>
<td></td>
</tr>
</tbody>
</table>

The data are provided as abdata.in7/abdata.bn7. Arellano and Bond (1991) consider dynamic employment equations of the form

\[ n_{i,t} = \alpha_1 n_{i,t-1} + \alpha_2 n_{i,t-2} + \beta' (L) x_{i,t} + \lambda_t + \eta_i + v_{i,t}, \]  

(6.1)

where \( x'_{i,t} = (w_{it}, k_{it}, y_{it}) \). Time dummies are included to capture aggregate demand shocks.

6.2 Data organization

The layout of a panel data set was introduced in §5.2. This section elaborates.

Note that we generally use the word individual to denote the cross-sectional unit (which could be individual, firm, country, etc.).
Because PcGive needs to recognize the structure of the data, and to allow for unbalanced panels, the data must be organized prior to analysis:

- Each variable is in a column.
- the columns are ordered by individual, and within individual by time.
- Each row refers to the same time period.
- The sample for each individual should be contiguous.
- a column with individual index is recommended (when omitted, the year variable is used to create such an index).
- There must be one column which denotes the year. This does not have to be the first column.
- An optional period column may be present when the data is not annual (i.e., the frequency is not one).
- An optional group column may be present to facilitate creating group dummies.

For example:

<table>
<thead>
<tr>
<th>year</th>
<th>index</th>
<th>employment</th>
<th>wages</th>
<th>individual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>1</td>
<td>110</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>1972</td>
<td>1</td>
<td>130</td>
<td>11</td>
<td>&quot;</td>
</tr>
<tr>
<td>1973</td>
<td>1</td>
<td>140</td>
<td>12</td>
<td>&quot;</td>
</tr>
<tr>
<td>1974</td>
<td>1</td>
<td>130</td>
<td>14</td>
<td>&quot;</td>
</tr>
<tr>
<td>1970</td>
<td>2</td>
<td>540</td>
<td>12</td>
<td>individual 2</td>
</tr>
<tr>
<td>1971</td>
<td>2</td>
<td>520</td>
<td>13</td>
<td>&quot;</td>
</tr>
<tr>
<td>1972</td>
<td>2</td>
<td>510</td>
<td>14</td>
<td>&quot;</td>
</tr>
<tr>
<td>1973</td>
<td>2</td>
<td>510</td>
<td>14</td>
<td>&quot;</td>
</tr>
</tbody>
</table>

Once the data set has the appropriate format, PcGive can determine where individuals start and finish by differencing the index variable. When the index variable is not specified, PcGive will use the year variable instead. This is done by differencing the year variable, and will exclude subsequent individuals where the year of the last observation on an individual is immediately after the year of the first observation on the next individual (which cannot happen with balanced panels, and may be a rare situation in unbalanced panels).

### 6.3 One-step GMM estimation

We start by replicating Table 4 column (a1) from Arellano and Bond (1991, p. 290). The lag structure of $x_{i,t}$ is one lag of $w_{it}$ and two lags on $k_{it}$ and $y_{it}$. The instrument matrix has the form:
The first two columns are normal instruments, the remainder are GMM-type instruments.

To formulate model (6.1) with instruments (6.2), first select Models for panel and then Dynamic Panel Methods using PcGive. Next, formulate the model as described above, not forgetting to add the Year variable. In the panel data package, instruments must be marked as such explicitly. So, select all the variables \( w \) to \( y_{s,2} \) in the model, and click on I: Instrument and X: Regressor. The dialog should look like:

The difference between the two types of instruments is as follows: I: Instrument will be differenced when the model is estimated in differences, whereas L: Level instrument would use them unchanged.
Click on OK, to move to the Functions dialog. If no GMM-type instruments are required, just leave this empty and press OK. To add the GMM-type instruments according to (6.2) select $n$ in the database, set $\text{Lag1}$ to 2 and $\text{Lag2}$ to 99, then press Add to see. The final specification for the GMM instruments is:

In the model settings add a constant and time dummies to the model, select Differences to estimate the model in first differences, as shown on the next page.

Scroll down in this dialog, and switch Use robust standard errors on (this was switched off in the previous chapter). Press OK and accept the default of 1-step estimation.

---

1Here, 99 is used as an upper limit, so that not more than 99 lags are used. In this case the actual upper limit is lower, but in general when $T$ gets larger, the large number of instruments can become a problem.
The output is:

DPD (1) Modelling n by 1-step

The dataset is: \.\OxMetrics7\data\abdata.in7

<table>
<thead>
<tr>
<th>1-step estimation using DPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
</tr>
<tr>
<td>Dn(-1)</td>
</tr>
<tr>
<td>Dn(-2)</td>
</tr>
<tr>
<td>Dw</td>
</tr>
<tr>
<td>Dw(-1)</td>
</tr>
<tr>
<td>Dk</td>
</tr>
<tr>
<td>Dk(-1)</td>
</tr>
<tr>
<td>Dk(-2)</td>
</tr>
<tr>
<td>Dys</td>
</tr>
<tr>
<td>Dys(-1)</td>
</tr>
<tr>
<td>Dys(-2)</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>T1980</td>
</tr>
<tr>
<td>T1981</td>
</tr>
</tbody>
</table>


\begin{verbatim}
Chapter 6 Tutorial on Dynamic Panel Data Modelling

T1982  -0.0248388  0.02027  -1.23  0.221
T1983  -0.00381603  0.02194  -0.174  0.862
T1984   0.00406269  0.02190  0.186  0.853

sigma   0.1239809  sigma^2   0.01537128
sigma levels   0.08766777
RSS  9.145908871  TSS  12.599978399
no. of observations  611  no. of parameters  16
Using robust standard errors

Transformation used:  
first differences
Transformed instruments:  w  w(-1)  k  k(-1)  k(-2)  ys  ys(-1)  ys(-2)
Level instruments:  Dummies  Gmm("n",2,99)

constant:  yes  time dummies:  5
number of individuals  140 (derived from year)
longest time series  6 [1979 - 1984]
shortest time series  4 (unbalanced panel)

Wald (joint):  Chi^2(10) = 408.3 [0.000] **
Wald (dummy):  Chi^2(6) = 11.58 [0.072]
Wald (time):  Chi^2(6) = 11.58 [0.072]
Sargan test:  Chi^2(25) = 65.82 [0.000] **
AR(1) test:  N(0,1) = -3.600 [0.000] **
AR(2) test:  N(0,1) = -0.5160 [0.606]

The standard errors and tests are based on the robust variance matrix, which is used by default (but can be switched off in the Model Settings). The AR(2) test is listed as \textit{m}_2 in Arellano and Bond (1991). Wald (joint) tests the significance on all regressors except the dummies, Wald (dummy) tests the significance of all dummies, and Wald (time) the significance of the time dummies and the constant (so the last two are identical).

6.4 Two-step GMM estimation

Two-step estimation is only different in the choice of estimation method. To replicate Arellano and Bond (1991, Table 4(b)), delete \( k_{i,t-1}, k_{i,t-2} \), and \( y_{i,t-2} \) from the model. Keep the model unchanged otherwise, selecting One and two-step estimation, which yields for the 2-step estimates (omitting the dummies from the output):

\begin{verbatim}
---- 2-step estimation using DPD ----
   Coefficient  Std.Error  t-value  t-prob
Dn(-1)    0.474151  0.1854  2.56  0.011
Dn(-2)  -0.0529675  0.05175  -1.02  0.306
Dw    -0.513205  0.1456  -3.53  0.000
Dw(-1)   0.224640  0.1419  1.58  0.114
Dk     0.292723  0.06263  4.67  0.000
Dys    0.609775  0.1563  3.90  0.000
Dys(-1) -0.446373  0.2173  -2.05  0.040

sigma   0.116243  sigma^2   0.01351243
sigma levels   0.08219621
RSS  8.0804358435  TSS  12.599978399
no. of observations  611  no. of parameters  13
\end{verbatim}
\end{verbatim}
Using robust standard errors

Transformation used: first differences
Transformed instruments: \( w, w(-1), k, ys, ys(-1) \)
Level instruments: Dummies Gmm(n,2,99)

Wald (joint): \( \text{Chi}^2(7) = 142.0 \ [0.000] \) **
Wald (dummy): \( \text{Chi}^2(6) = 16.97 \ [0.009] \) **
Sargan test: \( \text{Chi}^2(25) = 30.11 \ [0.220] \)
AR(1) test: \( N(0,1) = -1.538 \ [0.124] \)
AR(2) test: \( N(0,1) = -0.2797 \ [0.780] \)

The main differences with Table 4 column (b) of Arellano and Bond (1991) is in the reported standard errors. It was found that the estimated asymptotic variance matrix results in standard errors that are severely downward biased. Windmeijer (2000) derives a small-sample correction which is implemented in PcGive. It is used when robust standard errors are selected, as is the case here. Consequently, all derived test statistics are also different.

6.5 IV estimation

In Table 5, Arellano and Bond (1991) report the outcomes from other estimators, namely two Anderson-Hsiao type estimators, OLS and within-groups. These can all be estimated in PcGive. The first model is:

\[ \Delta n_{i,t-1} \]

Here, \( \Delta n_{i,t-1} \) is treated as endogenous, and \( \Delta n_{i,t-3} \) added as additional instrument. As before, the model includes time dummies and is estimated in differences, but it has no GMM-type instruments. The results from 1-step estimation match those in Table 5(e) (op. cit.).

The second Anderson-Hsiao type estimator has \( n_{i,t-3} \) as additional instrument instead of \( \Delta n_{i,t-3} \). Therefore, mark this variable as L: Level instrument. This ensures that the untransformed variable is used as additional instrument. Again, 1-step estimation reproduces Table 5(f) (op. cit.).

Arellano and Bond (1991) implemented the within-group estimator as OLS after applying the orthogonal deviations transformation. In unbalanced panels this is slightly
different from estimation in deviations from the means of each individual (but asymptotically identical). A comparison using PcGive confirms that the difference is small.

### 6.6 Combined GMM estimation

PcGive also allows estimation using the combined GMM-SYS estimator proposed in Arellano and Bover (1995) and Blundell and Bond (1998). In this type of estimation, the level equations are stacked on top of the transformed equations.

Blundell and Bond (1998) consider a dynamic employment equation of the form

\[ n_{i,t} = \alpha_1 n_{i,t-1} + \beta_1 w_{i,t} + \beta_2 w_{i,t-1} + \gamma_1 k_{i,t} + \gamma_2 k_{i,t-1} + \eta_i + \nu_{i,t}, \]

using the same data set as Arellano and Bond (1991):

The GMM-type instruments for the differenced equations are

\[ \text{diag}(n_{i,1} \cdots n_{i,t-2}, w_{i,1} \cdots w_{i,t-2}, k_{i,1} \cdots k_{i,t-2}), \]

which is formulated as:

The GMM-style instruments in the levels equation are the lagged differences:

\[ \text{diag}(\Delta n_{i,t-1}, \Delta w_{i,t-1}, \Delta k_{i,t-1}). \]
These are added in the same dialog. Note that now the first argument is the lag length, and the second whether to use differences (1) or not (0):

The presence of GMM-level instruments will automatically result in combined estimation. At the Model Settings, add time dummies, and estimate in differences by 1 and 2 step estimation. The 2-step results are:

```
---- 2-step estimation using DPD ----

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dn(-1)</td>
<td>0.872881</td>
<td>0.04528</td>
<td>19.3</td>
</tr>
<tr>
<td>Dw</td>
<td>-0.779745</td>
<td>0.1166</td>
<td>-6.69</td>
</tr>
<tr>
<td>Dw(-1)</td>
<td>0.526803</td>
<td>0.1621</td>
<td>3.25</td>
</tr>
<tr>
<td>Dk</td>
<td>0.470077</td>
<td>0.07986</td>
<td>5.89</td>
</tr>
<tr>
<td>Dk(-1)</td>
<td>-0.357608</td>
<td>0.08003</td>
<td>-4.47</td>
</tr>
<tr>
<td>Constant</td>
<td>0.948489</td>
<td>0.3776</td>
<td>2.51</td>
</tr>
<tr>
<td>T1978</td>
<td>0.00580177</td>
<td>0.01971</td>
<td>0.294</td>
</tr>
<tr>
<td>T1979</td>
<td>0.0188976</td>
<td>0.02277</td>
<td>0.830</td>
</tr>
<tr>
<td>T1980</td>
<td>0.00281961</td>
<td>0.02407</td>
<td>0.117</td>
</tr>
<tr>
<td>T1981</td>
<td>-0.0200226</td>
<td>0.02744</td>
<td>-0.730</td>
</tr>
<tr>
<td>T1982</td>
<td>0.0152802</td>
<td>0.02331</td>
<td>0.656</td>
</tr>
<tr>
<td>T1983</td>
<td>0.0317310</td>
<td>0.02350</td>
<td>1.35</td>
</tr>
<tr>
<td>T1984</td>
<td>0.0224205</td>
<td>0.03107</td>
<td>0.722</td>
</tr>
</tbody>
</table>

sigma levels 0.1276452 sigma^2 levels 0.01629331
RSS 14.305523988 TSS 1601.0425015
no. of observations 891 no. of parameters 13
Using finite sample corrected standard errors

Transformation used: first differences
GMM-SYS estimation combines transformed and level equations

Instruments for transformed equation:
Level instruments: Gmm(n,2,99) Gmm(w,2,99) Gmm(k,2,99)

Instruments for level equations:
Level instruments: Dummies GmmLevel(n,1,1) GmmLevel(w,1,1)
GmmLevel(k,1,1)

constant: yes time dummies: 7
number of individuals 140 (derived from year)
longest time series 7 [1978 - 1984]
shortest time series 5 (unbalanced panel)

Wald (joint): $\chi^2(5) = 3378. [0.000]$ **
Wald (dummy): $\chi^2(8) = 28.51 [0.000]$ **
Wald (time): $\chi^2(7) = 16.00 [0.025]$ *
Sargan test: $\chi^2(100) = 111.6 [0.201]$ 
AR(1) test: $N(0,1) = -5.807 [0.000]$ ** 
AR(2) test: $N(0,1) = -0.1471 [0.883]$ 

These results do not exactly match the last column of Blundell and Bond (1998, Table 4) for several reasons:

1. We report 2-step results instead of 1-step results.
2. The 2-step results in PcGive now use the estimated variance matrix incorporating the small-sample correction of Windmeijer (2000).
3. In Blundell and Bond (1998), the differenced dummies were used as instruments in the transformed equations, and the $H_i$ matrix was set to the identity matrix.
Chapter 7

Panel Data Implementation Details

This section gives a summary of the statistical output of PcGive, giving the formulae which are used in the computations.

7.1 Transformations

- none
  \[ x_{it}^* = x_{it}, \quad t = 1, \ldots, T_i. \]

- first differences
  \[ x_{it}^* = \Delta x_{it} = x_{it} - x_{i,t-1}, \quad t = 2, \ldots, T_i, \]
  \[ x_{i1}^* = \Delta x_{i1} = 0. \]

- time means:
  \[ \bar{x}_i = \frac{1}{T_i} \sum_{s=1}^{T_i} x_{is}. \]

- deviations from time means
  \[ x_{it}^* = x_{it} - \bar{x}_i, \quad t = 1, \ldots, T. \]

- orthogonal deviations
  \[ x_{it}^o = \left( x_{it} - \frac{1}{T_i - t} \sum_{s=t+2}^{T_i} x_{is} \right) \left( \frac{T_i - t}{T_i - t + 1} \right)^{1/2}, \quad t = 2, \ldots, T_i - 1. \]

  The orthogonal deviations are stored with one lag, so that the first observation is lost instead of the last (this brings it in line with first differencing):
  \[ x_{it}^* = x_{i,t-1}^o, \quad t = 2, \ldots, T_i. \]
  \[ x_{i1}^* = 0. \]

- GLS deviations:
  \[ x_{it}^* = x_{it} - \theta_i \bar{x}_i, \quad t = 1, \ldots, T_i, \]
  where the choice of \( \theta_i \) determines the method, as discussed in the next section.
7.2 Static panel-data estimation

The starting point is (4.2):
\[ y = W\beta + v. \]

The total number of observations is \( O = NT \). In an unbalanced panel, different individuals may have a different number of observations, \( T_i \). In that case the total number of observations equals \( O = \sum T_i \). Baltagi (1995) reviews the standard estimation methods used in this setting:

- The OLS estimates are:
  \[ \hat{\beta} = (W'W)^{-1}W'y, \]
  \[ \hat{\sigma}^2_v = \hat{\nu}'\hat{\nu} / (n - p), \]
  \[ \hat{\nu} = y - W\hat{\beta}. \]

Here \( W \) has \( k \) columns, containing all specified regressors including dummies, so \( p = k \).

- The LSDV (least squares dummy variables) estimates use individual dummies.
- The within estimates replace \( y \) and \( X \) by deviations from time means.
- The between estimates replace \( y \) and \( W \) by the individual means.
- The feasible GLS estimates replace \( y \) and \( W \) by deviations from weighted time means. The outcome depends on the choice of \( \theta \), which in PcGive is set by specifying \( \sigma^2_v \) and \( \sigma^2_\eta \):
  \[ \theta_i = 1 - \frac{\sigma_v}{\sigma_i}, \quad \sigma_i^2 = \sigma_v^2 + T_i \sigma_\eta^2. \]

When OLS residuals \( u = \hat{\nu}_{OLS} \) are used, the GLS estimator can be based on:

\[ \hat{\sigma}^2_v = \frac{\sum_{i=1}^N \sum_{t=1}^{T_i} (u_{it} - \bar{u}_i)^2}{O - N}, \]

\[ \hat{\sigma}_i^2 = \sigma_0^2 = \frac{\sum_{i=1}^N T_i \bar{u}_i^2}{N - 1}. \]

PcGive computes \( \theta_i \) from \( \sigma_v^2 \) and \( \sigma_\eta^2 \) with the latter derived from \( \hat{\sigma}_0^2 \) as:

\[ \hat{\sigma}_\eta^2 = (\hat{\sigma}_0^2 - \hat{\sigma}_v^2) \frac{N}{O}. \]

In a balanced panel: \( N/O = 1/T \). This is not optimal in an unbalanced panel, but seems reasonable.

The standard feasible GLS estimator uses the between and within estimates:

\[ \sigma_v^2 = \hat{\sigma}_{\text{within}}, \]
\[ \sigma_i^2 = T_i \hat{\sigma}_{\text{between}}. \]
7.3 Dynamic panel data estimation

For convenience, \( \theta_i \) is specified using three variance components \( \sigma_v^2, \sigma_a^2, \) and \( \sigma_\eta^2 \):

\[
\theta_i = 1 - \left( \frac{\sigma_v^2}{\sigma_a^2 + T_i \sigma_\eta^2} \right)^{1/2}.
\]

(7.1)

<table>
<thead>
<tr>
<th>OLS based GLS</th>
<th>( \hat{\sigma}_v^2 )</th>
<th>( \hat{\sigma}_a^2 )</th>
<th>( \hat{\sigma}_\eta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>between/within based GLS</td>
<td>( \hat{\sigma}_v^2 )</td>
<td>( \hat{\sigma}_v^2 )</td>
<td>( \hat{\sigma}_\eta^2 )</td>
</tr>
</tbody>
</table>

- The *maximum likelihood* estimates obtain \( \theta \) by iterating the GLS procedure. The concentrated likelihood is:

\[
\ell_c(\tau; \hat{\beta}(\tau), \hat{\sigma}_v^2(\tau))/O = c - 0.5 \log(\hat{\sigma}_v^2) - 0.5 \sum_{i=1}^{N} \log(1 + T_i \tau)/O,
\]

(7.2)

where \( \tau = \sigma_\eta^2/\sigma_v^2 \), so that:

\[
\theta_i = 1 - (1 + T_i \tau)^{-1/2}.
\]

To summarize the implementation in PcGive:

<table>
<thead>
<tr>
<th>number of observations, ( n )</th>
<th>degrees of freedom lost in estimation, ( p )</th>
<th>transforms dummies</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS (no transformation)</td>
<td>( O )</td>
<td>( k )</td>
</tr>
<tr>
<td>within</td>
<td>( O )</td>
<td>( k + N )</td>
</tr>
<tr>
<td>between</td>
<td>( N )</td>
<td>( k )</td>
</tr>
<tr>
<td>GLS, ML</td>
<td>( O )</td>
<td>( k )</td>
</tr>
</tbody>
</table>

It is important to note that the within transformations are only applied to \( X \), and not to the dummies \( D \). In the within estimator, individual dummies are redundant after subtracting the individual means. In the between estimator, time dummies are irrelevant, and individual dummies create a perfect fit. In GLS and ML, the individual means are subtracted with weights.

### 7.3 Dynamic panel data estimation

The single equation dynamic panel model can be written as:

\[
y_{it} = \sum_{s=1}^{m} \alpha_{i} y_{i,t-s} + x'_{it} \gamma + \lambda_t + \eta_i + v_{it}, \quad t = 0, \ldots, T_i - 1, \quad i = 0, \ldots, N - 1.
\]

As before, \( \lambda_t \) and \( \eta_i \) are time and individual specific effects and \( x_{it} \) is a \( k^* \) vector of explanatory variables. It is assumed that allowance for lagged observations has been
made, so observations on \( y_{i,-s}, \ldots, y_{i,-1} \) are available. The total number of observations available for estimation equals \( O = \sum T_i \).

Stacking the data for an individual according to time: and using \( D \) for the matrix with dummies:

\[
y_i = X_i \gamma + D_i \delta + v_i, \quad i = 0, \ldots, N - 1.
\]

In formulating the model, we distinguish the following types of variables:

- \( y \): dependent variables,
- \( X \): regressors, including lagged dependent variables,
- \( D \): dummy variables,
- \( I \): normal instruments,
- \( L \): ‘level’ instruments,
- \( G \): GMM-style instruments.

The \( G_i \) are described under the \( \text{Gmm} \) function. From these variables, the dependent variable, regressor and instrument matrices are formed as follows:

<table>
<thead>
<tr>
<th>( q_i )</th>
<th>dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_i^* )</td>
<td>( y_i \times 1 )</td>
</tr>
<tr>
<td>( X_i^* : D_i^* )</td>
<td>( X_i \times k )</td>
</tr>
<tr>
<td>( G_i^* : I_i^* : D_i^* : L_i )</td>
<td>( G_i \times z )</td>
</tr>
</tbody>
</table>

where the \( * \) denotes some transformation which can be chosen from \( \S 7.1 \). This also affects the degrees of freedom, as shown in Table 7.1. The estimators and tests are described in Arellano and Bond (1991), and have the form:

\[
\begin{align*}
M &= \left( \sum_i W_i'Z_i \right) A_N \left( \sum_i Z_i'W_i \right), \\
A_N^{-1} &= \left( \sum_i Z_i'H_iZ_i \right)^{-1}, \\
\hat{\beta} &= M^{-1} \left( \sum_i W_i'Z_i \right) A_N \left( \sum_i Z_i'q_i \right), \\
\hat{\sigma}_u^2 &= \hat{u}'\hat{u} / (n - p), \\
\hat{u} &= q - W\hat{\beta}.
\end{align*}
\]

\[ (7.3) \]

**Table 7.1**  Transformation and the treatment of dummy variables

<table>
<thead>
<tr>
<th>transformation</th>
<th>number of observations, ( n )</th>
<th>degrees of freedom lost in estimation, ( p )</th>
<th>transforms dummies</th>
</tr>
</thead>
<tbody>
<tr>
<td>differences</td>
<td>( O - N )</td>
<td>( k )</td>
<td>( D^* = D ) (optional: ( D^* ))</td>
</tr>
<tr>
<td>orthogonal deviations</td>
<td>( O - N )</td>
<td>( k )</td>
<td>( D^* = D ) (optional: ( D^* ))</td>
</tr>
<tr>
<td>none</td>
<td>( O )</td>
<td>( k )</td>
<td>no: ( D^* = D )</td>
</tr>
<tr>
<td>within</td>
<td>( O )</td>
<td>( k + N )</td>
<td>no: ( D^* = D^* )</td>
</tr>
<tr>
<td>between</td>
<td>( N )</td>
<td>( k )</td>
<td>yes: ( D^* = D^* )</td>
</tr>
<tr>
<td>GLS</td>
<td>( O )</td>
<td>( k )</td>
<td>yes: ( D^* = D^* )</td>
</tr>
</tbody>
</table>
When the option to concentrate out dummies is used, $y^*, X^*, I^*$, and $L$ are replaced by the residuals from regressing on the set of dummies $D^*$. Subsequently, the dummies are omitted from further computations.

In one-step estimation, $H_i$ is the identity matrix, except when estimating in first differences:

$$
H_{1,i} = I_{T_i}, \text{ except: } H_{1,i}^{\text{diff}} = \begin{pmatrix}
1 & -1/2 & 0 & \cdots & 0 \\
-1/2 & 1 & -1/2 & 0 \\
0 & -1/2 & 1 & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \cdots & 1
\end{pmatrix}.
$$

In two-step estimation $H_i$ is based on the previous step residuals:

<table>
<thead>
<tr>
<th>$H_i$</th>
<th>$M$</th>
<th>$A_N$</th>
<th>$\hat{u}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>one-step $H_{1,i}$, see (7.4)</td>
<td>$M_1$</td>
<td>$A_{1,N}$</td>
<td>$\hat{u}<em>{1,i}$ = $\hat{v}</em>{1,i}^*$</td>
</tr>
<tr>
<td>two step $H_{2,i}$</td>
<td>$\hat{v}<em>{1,i}^* \hat{v}</em>{1,i}^\prime$</td>
<td>$M_2$</td>
<td>$A_{2,N}$</td>
</tr>
</tbody>
</table>

with a subscript added to distinguish between one and two step estimation.

The output reports the residual and total sum of squares:

$$
\text{RSS} = \hat{u}' \hat{u},
$$

$$
\text{TSS} = \hat{q}' \hat{q} - (\hat{q}' \hat{q})^2 / O.
$$

So the TSS is in deviation from the mean.

The one-step estimated variance of the coefficients is:

$$
\hat{V}_1 [\hat{\beta}] = \hat{\sigma}_{1,u}^2 M_1^{-1}.
$$

If selected, robust standard errors after one-step estimation are computed as:

$$
\hat{V}_{1r} [\hat{\beta}] = M_1^{-1} \left( \sum_i W_i' Z_i \right) A_{1,N} A_{2,N}^{-1} A_{1,N} \left( \sum_i Z_i' W_i \right) M_1^{-1}.
$$

When there are no instruments, (7.5) simplifies to:

$$
(W'W)^{-1} \left( \sum_{i=1}^N W_i' \hat{u}_i \hat{u}_i' W_i \right) (W'W)^{-1},
$$

which is the robust variance estimator proposed in Arellano (1987).\footnote{This robust estimator is not identical to the standard White (1980) estimator for OLS, which uses as the central matrix: $\sum_{i=1}^{NT} w_i' \hat{u}_i \hat{u}_i' w_i$.} The two-step asymptotic variance matrix is:

$$
\hat{V}_2 [\hat{\beta}] = M_2^{-1}.
$$
This variance estimator can be severely downward biased in small samples, and is therefore only reported when robust standard errors is switched off. The preferred solution (and the default) is to use the robust variance for two-step estimation, $\hat{V}_{2r}[\hat{\beta}]$, which uses the small sample correction as derived by Windmeijer (2000).

The AR test for order $m$ is computed as:

$$AR(m) = \frac{d_0}{(d_1 + d_2 + d_3)^{1/2}}. \tag{7.6}$$

Using $w_i$ for the residuals lagged $m$ periods (substituting zero for missing lags):

$$w_{it} = u_{i,t-m} \text{ for } t = m, \ldots, T_i, \text{ and } w_{it} = 0 \text{ for } t < m.$$

The $d_i$ are defined as:

$$\begin{align*}
    d_0 &= \sum_i w_i' u_i, \\
    d_1 &= \sum_i w_i' H_i w_i, \\
    d_2 &= -2 \left( \sum_i w_i' W_i \right) M^{-1} \left( \sum_i W_i' Z_i \right) A_N \left( \sum_i Z_i' H_i w_i \right), \\
    d_3 &= \left( \sum_i w_i' W_i \right) \hat{V}[\hat{\beta}] \left( \sum_i W_i' w_i \right).
\end{align*} \tag{7.7}$$

The components are used as the notation suggests, except for $H_i$ in one-step estimation:

<table>
<thead>
<tr>
<th></th>
<th>$H_i$</th>
<th>$M$</th>
<th>$A_N$</th>
<th>$u_i$</th>
<th>$\hat{V}[\hat{\beta}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>one step</td>
<td>$\hat{\sigma}^2_{1,i} H_{1,i}$</td>
<td>$M_1$</td>
<td>$A_{1,N}$</td>
<td>$\tilde{v}_{1,i}$</td>
<td>$\hat{V}_1[\hat{\beta}]$</td>
</tr>
<tr>
<td>one step, robust</td>
<td>$H_{2,i}$</td>
<td>$M_1$</td>
<td>$A_{1,N}$</td>
<td>$\tilde{v}_{1r,i}$</td>
<td>$\hat{V}_{1r}[\hat{\beta}]$</td>
</tr>
<tr>
<td>two step</td>
<td>$H_{2,i}$</td>
<td>$M_2$</td>
<td>$A_{2,N}$</td>
<td>$\tilde{v}_{2,i}$</td>
<td>$\hat{V}_2[\hat{\beta}]$</td>
</tr>
<tr>
<td>two step, robust</td>
<td>$H_{2,i}$</td>
<td>$M_2$</td>
<td>$A_{2,N}$</td>
<td>$\tilde{v}_{2r,i}$</td>
<td>$\hat{V}_{2r}[\hat{\beta}]$</td>
</tr>
</tbody>
</table>

After estimation in orthogonal deviations, the AR test (7.6) is based on the residuals from the differenced model as follows (the superscript $\Delta$ refers to the regressors and residuals from the model in differenced form):

$$\begin{align*}
    d_0 &= \sum_i w_i^\Delta' u_i^\Delta, \\
    d_1 &= \sum_i w_i^\Delta' H_i^\Delta w_i^\Delta, \\
    d_2 &= -2 \left( \sum_i w_i^\Delta' W_i^\Delta \right) M^{-1} \left( \sum_i W_i' Z_i \right) A_N \left( \sum_i Z_i' \Psi_i \right), \\
    d_3 &= \left( \sum_i w_i^\Delta' W_i^\Delta \right) \hat{V}[\hat{\beta}] \left( \sum_i W_i^\Delta' w_i^\Delta \right).
\end{align*} \tag{7.8}$$

$H_i^\Delta$ is $H_{1,i}^{\text{diff}}$ after one-step estimation, and $u_i^\Delta u_i^\Delta$ otherwise. $\Psi_i$ equals

$$\begin{pmatrix}
    \sqrt{(T_i + 1)/T_i} & 0 & 0 & \cdots & 0 \\
    \sqrt{(T_i - 1)/(T_i - 2)} & \sqrt{T_i/(T_i - 1)} & 0 & \cdots & 0 \\
    \vdots & \vdots & \ddots & \cdots & \vdots \\
    0 & 0 & 0 & \sqrt{1/2} & \sqrt{2/1}
\end{pmatrix}$$

$w_i$ after one-step estimation, and $u_i^\Delta u_i^\Delta$ after one-step robust or two-step estimation.
The Sargan test with \( z - k \) degrees of freedom is after one-step estimation:

\[
S_1 = \left( \sum_i \hat{v}_{1,i}^* Z_i \right) A_{1,N} \left( \sum_i Z_i' \hat{v}_{1,i}^* \right) \sigma_{1,1}^{-2},
\]

and after two-step estimation (see 4.4 and equation (10) in Arellano and Bond, 1991):

\[
S = \left( \sum_i \hat{v}_{2,i}^* Z_i \right) A_{2,N} \left( \sum_i Z_i' \hat{v}_{2,i}^* \right).
\]

Three Wald tests are routinely reported to test the significance of groups of variables:

- Wald (joint): on \( W \), all regressors except dummies,
- Wald (dummy): on \( D \), all dummies (including constant),
- Wald (time): on time dummies (including the constant in the differenced/deviations model).

### 7.4 Dynamic panel data, combined estimation

In combined estimation, GMM-SYS, the levels equation is used jointly with the transformed equation, see Blundell and Bond (1998). The IV estimation of (7.3) still applies, but the data are organized as follows:

<table>
<thead>
<tr>
<th>Dimension</th>
<th>( q_i )</th>
<th>( W_i )</th>
<th>( Z_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( q_i )</td>
<td>( q_i^+ )</td>
<td>( W_i^* )</td>
<td>( Z_i^+ )</td>
</tr>
<tr>
<td>( q_i^+ )</td>
<td>( y_i^* )</td>
<td>( X_i^* )</td>
<td>( G_i^+ )</td>
</tr>
<tr>
<td>( y_i^* )</td>
<td>( y_i )</td>
<td>( X_i )</td>
<td>( D_i )</td>
</tr>
<tr>
<td>( y_i )</td>
<td>( y_i )</td>
<td>( X_i )</td>
<td>( D_i )</td>
</tr>
<tr>
<td>( X_i^* )</td>
<td>( X_i )</td>
<td>( D_i )</td>
<td></td>
</tr>
<tr>
<td>( X_i )</td>
<td>( X_i )</td>
<td>( D_i )</td>
<td></td>
</tr>
<tr>
<td>( G_i^+ )</td>
<td>( G_i^+ )</td>
<td>( D_i )</td>
<td></td>
</tr>
<tr>
<td>( G_i^+ )</td>
<td>( G_i^+ )</td>
<td>( D_i )</td>
<td></td>
</tr>
<tr>
<td>( L_i )</td>
<td>( L_i )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( G_i^+ \) are the GMM-style instruments for the levels equation, described under the GmmLevel function.² The dummies in the transformed equation are always transformed. When using differences or deviations, the effective number of observations in the transformed equations is as before \((O - N)\), whereas the levels equations have \( O \) observations. Internally, both the transformed and levels equation \( i \) has \( T_i \) equations, but, when using differences/deviations, the first observation in the transformed equation is set to zero, resulting in \( T_i - 1 \) effective observations.

When the option to concentrate out dummies is used, \( y^*, X^*, I^*, \) and \( L \) are replaced by the residuals from regressing on the set of dummies \( D^* \), and \( y \) and \( X \) on \( D \). Note that there is one time dummy extra in combined estimation compared to normal estimation when using differences or deviations.

² The code used to estimate Blundell and Bond (1998) had \( Z_i^* = (G_i \ 0 \ 1^*_i \ D_i^* \ L_i) \).
After estimation, there are two sets of residuals:

\[ \hat{u}^* = q^* - W^*\hat{\beta} \] (transformed equation),
\[ \hat{u}^+ = q^+ - W^+\hat{\beta} \] (levels equation).

The reported residual variance is based on the levels residuals: \( \hat{\sigma}^2 u^+ = \hat{u}^+\hat{u}^+/(n-p) \). Here \( n = O = \sum_{i=1}^{N} T_i \) and \( p \) are as before. The reported RSS and TSS are also based on the levels, using \( \hat{u}^+ \) and \( q^+ \).

**In one-step estimation**, \( H_i \) is as in (7.4) in the transformed equations, and the identity matrix in the level equations (\( 1/2 I \) when using differences).\(^3\) Only robust standard errors are reported after one-step estimation.

**In two-step estimation**, \( H_i \) is based on the previous step residuals \( \hat{u}'_i = (\hat{u}^* : \hat{u}^+). \)

The AR test (7.6) is based on the residuals from the transformed equation:

\[
\begin{align*}
    d_0 &= \sum_i w_i' u_i^*, \\
    d_1 &= \sum_i w_i' u_i^* u_i^* w_i^*, \\
    d_2 &= -2 (\sum_i w_i' W_i^*) M^{-1} (\sum_i W_i' Z_i) A_N (\sum_i Z_i' u_i u_i^* w_i^*), \\
    d_3 &= (\sum_i w_i' W_i^*) \hat{V} \left[ \hat{\beta} \right] (\sum_i W_i' w_i^*).
\end{align*}
\]

After orthogonal deviations, the * is replaced by \( \Delta \) as in (7.8).

### 7.5 Panel batch commands

\[ \text{estimate("method"=OLS, year1=\text{-1}, period1=\text{0}, year2=\text{-1}, period2=\text{0});} \]

Estimate a system. Note that for panel data all sample arguments are ignored: only the estimation method is used.

The method argument is one of:

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;1-step&quot;</td>
<td>1-step estimation,</td>
</tr>
<tr>
<td>&quot;2-step&quot;</td>
<td>1 and 2-step estimation,</td>
</tr>
<tr>
<td>&quot;OLS&quot;</td>
<td>OLS,</td>
</tr>
<tr>
<td>&quot;LSDV&quot;</td>
<td>Least-squares dummy variables,</td>
</tr>
<tr>
<td>&quot;Within&quot;</td>
<td>within-groups estimation,</td>
</tr>
<tr>
<td>&quot;Between&quot;</td>
<td>between-groups estimation,</td>
</tr>
<tr>
<td>&quot;GLS(w/b)&quot;</td>
<td>GLS (using within/between),</td>
</tr>
<tr>
<td>&quot;GLS&quot;</td>
<td>GLS (OLS based),</td>
</tr>
<tr>
<td>&quot;ML&quot;</td>
<td>Maximum likelihood.</td>
</tr>
</tbody>
</table>

\[ \text{Gmm("variable", lag1, lag2);} \]

Specifies a GMM type instrument.

---

\(^3\)The code used to estimate in [Blundell and Bond (1998)](https://www.jstor.org/stable/2600384) had \( H_{1,i} = I_{T_i} \) also for estimation in differences.
GmmLevel("variable", lag, isdiff);

Specifies a level GMM type instrument for combined estimation. If the isdiff argument is 1 it is added in differences, if 0 in levels.

module("PcGive");

Starts the PcGive module. If PcGive is already running, this batch command is not required.

option("option", argument);

<table>
<thead>
<tr>
<th>option</th>
<th>argument</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>robust</td>
<td>Use robust standard errors</td>
<td>0 for off, 1 on,</td>
</tr>
<tr>
<td>concentrate</td>
<td>Concentrate dummies</td>
<td>0 for off, 1 on,</td>
</tr>
<tr>
<td>transformdummy</td>
<td>Transform dummies (not exact with instruments)</td>
<td></td>
</tr>
<tr>
<td>printgmm</td>
<td>Print contents of GMM instruments</td>
<td>0 for off, 1 on,</td>
</tr>
<tr>
<td>maxit</td>
<td>maximum number of iterations</td>
<td>default: 1000,</td>
</tr>
<tr>
<td>print</td>
<td>print every # iteration</td>
<td>0: do not print,</td>
</tr>
<tr>
<td>compact</td>
<td>compact or extended output</td>
<td>0 for off, 1 on,</td>
</tr>
<tr>
<td>strong</td>
<td>strong convergence tolerance</td>
<td>default: 0.005,</td>
</tr>
<tr>
<td>weak</td>
<td>set weak convergence tolerance</td>
<td>default: 0.0001,</td>
</tr>
</tbody>
</table>

package("Panel", "package");

Use this command to select the correct Panel component (package) from PcGive:

```
package
"Static"
"Dynamic"
```

setdummy("constant", ...);

Lists the dummy variables to be added. The arguments are separated by a comma and can be: "constant", "time", "group", "timegroup", "individual".

settest(ispec, iar);

Sets the test options. The first argument is 0 or 1, and determines whether specification tests are printed. The second argument sets the order for the AR test; the default is 2.

settransform("transform");

Lists the desired transformation. The argument can be: "differences", "deviations", "within", "between".

store("name", "rename"="");

Use this command to store residuals, etc. into the database, the default name is used. Note that if the variable already exists, it is overwritten without warning. The name must be one of:

- residuals residuals
- resdiff differenced residuals (after combined estimation)
- reslevels levels residuals (after combined estimation)

The optional second argument replaces the default name. For exam-
store("residuals") stores the residuals under the name Residual; store("residuals", "xyz") stores them under the name xyz.

system { Y=...; X=...; I=...; L=...; R=...; N=...; G=...; }
  Specify the system, consisting of the following components:
  Y endogenous variable;
  X non-modelled X variable;
  I Instrument;
  L Level instrument.
  R Year variable.
  N Index variable.
  G Group variable.
  The variables listed are separated by commas, their base names (that is, name excluding lag length) must be in the database. If the variable names are not a valid token, the name must be enclosed in double quotes.
  Note that IV formulation is different here from the econometric modelling package:

<table>
<thead>
<tr>
<th></th>
<th>Econometric modelling</th>
<th>Panel data models</th>
</tr>
</thead>
<tbody>
<tr>
<td>dependent variable:</td>
<td>Y (first Y)</td>
<td>Y (just one)</td>
</tr>
<tr>
<td>endogenous regressors:</td>
<td>Y (remaining Y’s)</td>
<td>X</td>
</tr>
<tr>
<td>regressors:</td>
<td>Z</td>
<td>X, L or X, I</td>
</tr>
<tr>
<td>additional instruments:</td>
<td>A</td>
<td>L or I</td>
</tr>
</tbody>
</table>

\[\text{testlinres}\{\ldots\}\]
  Test for linear restrictions. The content is the matrix dimensions followed by the \((R : r)\) matrix.
Part IV

Volatility Models (GARCH)

with H. Peter Boswijk and Marius Ooms
Chapter 8

Introduction to Volatility Models

8.1 Introduction

The autoregressive conditional-heteroscedastic (ARCH, see Engle, 1982) and generalized ARCH (GARCH, see Bollerslev, 1986) models have found widespread application since their introduction. These models were designed to capture the volatility clustering which can be observed in macro-economic series such as inflation (Engle, 1982), or financial time-series such as exchange rates and stock market returns. Recent surveys of GARCH-type models include Bera and Higgins (1993), Bollerslev, Engle, and Nelson (1994), Shephard (1996), and Gourieroux (1997), also see the text books by Hamilton (1994), Mills (1999) and Tsay (2005).

The ARCH \( (q) \) model featured briefly in Volume I as part of non-linear modelling, and throughout in the form of the LM test for ARCH effects in the residuals of a regression model. The ARCH \( (q) \) model can be defined as:

\[
y_t = \mu + u_t, \quad \text{with} \quad u_t = \sqrt{h_t} \varepsilon_t, \quad \text{and} \quad h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j u_{t-j}^2.
\] (8.1)

Assuming \( \varepsilon_t \sim \text{IN}[0,1] \), identically and independently distributed for all \( t \), gives \( u_t|u_{t-1} \cdots u_{t-q} \sim N(0, h_t) \). So conditional on the past, the model is normal but heteroscedastic. The log-likelihood for observation \( t \) differs from the linear model with normal errors through the time-dependence of the conditional variance:

\[
\ell_t (\mu, \alpha|\mathcal{F}_{t-1}) = c - \frac{1}{2} \log (h_t) - \frac{1}{2} (\frac{y_t - \mu}{h_t})^2,
\]

where \( \mathcal{F}_{t-1} \) captures all the information up to \( t-1 \).

In contrast with stochastic volatility models (see Shephard, 1996), where the variance is modelled as an unobserved (latent) process, the model in (8.1) leads to an explicit expression for the log-likelihood: given values for \( \mu \) and \( \alpha = (\alpha_1, \ldots, \alpha_q)^t \), we can evaluate \( \tilde{u}_{t-q}, \ldots, \tilde{u}_{t-1} \) and therefore \( \tilde{h}_t \) and \( \ell_t \). This makes maximum likelihood
estimation of the model in (8.1), and its extensions discussed below, relatively easy in comparison with stochastic volatility models. There is a problem very early on in this recursion regarding the treatment of \( \hat{u}_t \) for \( t < 0 \) — this is discussed below. The first-order conditions for a maximum of the log-likelihood cannot be solved analytically in terms of the parameters; instead this requires the use of a numerical optimization routine.

More generally, the regression model with normal-GARCH\((p,q)\) errors is defined as:

\[
y_t = x_t'\zeta + u_t, \\
u_t = \varepsilon_t h_t^{1/2}, \quad \varepsilon_t|F_{t-1} \sim N[0,1], \\
h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \quad t = 1, \ldots, T.
\]  

(8.2)

The ARCH\((q)\) model corresponds to GARCH\((0,q)\). The GARCH\((p,q)\) model allows for a parsimonious parametrization of an ARCH\((\infty)\) structure: assuming that the roots of \( 1 - \beta(L) = 1 - \sum_{i=1}^{p} \beta_i L^i \) are outside the unit circle, and letting \( \alpha(L) = \sum_{i=1}^{q} \alpha_i L^i \), (8.2) implies

\[
h_t = \frac{\alpha_0}{1 - \beta(L)} + \frac{\alpha(L)}{1 - \beta(L)} u_t^2.
\]

The conditional variance \( h_t \) must be nonnegative, which requires that the coefficients of \( \delta(L) = \alpha(L)/(1 - \beta(L)) \) are nonnegative (and \( \alpha_0 \geq 0 \)). A sufficient condition for this is that \( \alpha_i \geq 0 \) and \( \beta_i \geq 0 \) for all \( i \) (however this is overly restrictive; see Nelson and Cao, 1992).

The equation for \( h_t \) in (8.2) can be written in ARMA form using \( \nu_t = u_t^2 - h_t = (\varepsilon_t^2 - 1) h_t \):

\[
u_t^2 = \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i) u_{t-i}^2 - \sum_{i=1}^{p} \beta_i \nu_{t-i} + \nu_t,
\]

(8.3)

where \( m = \max(p,q) \) and we set \( \beta_i = 0 \) for \( i > p \), and \( \alpha_i = 0 \) for \( i > q \). Note that \( E(\nu_t|F_{t-1}) = 0 \). It follows that \( u_t \) has a finite unconditional variance, given by \( E[u_t^2] = \alpha_0/(1 - \alpha(1) - \beta(1)) \), if and only if the roots of

\[1 - \alpha(L) - \beta(L) = 1 - \sum_{i=1}^{m} (\alpha_i + \beta_i) L^i\]

are all outside the unit circle. A necessary and (given nonnegativity of \( \alpha_i \) and \( \beta_i \)) sufficient condition for this is \( \alpha(1) + \beta(1) = \sum_{i=1}^{m} (\alpha_i + \beta_i) < 1 \). Under this condition, the process \( u_t \) is weakly stationary. Equation (8.3) also forms the basis for forecasting, because, for the GARCH\((1,1)\) model:

\[E[h_{T+H}|h_T] = (\alpha + \beta)^H h_T + \alpha_0 \left[ \sum_{j=0}^{H-1} (\alpha + \beta)^j \right].\]
In the first forecast, $\nu_{T-1}$ still features, but afterwards it is set to its expectation of zero. The default parameter space in PcGive is formulated in terms of the ARMA representation, with $\alpha_0 \geq 0$, $\alpha_i + \beta_i \geq 0$, and $\alpha(1) + \beta(1) < 1$.\(^1\)

When $\alpha(1)+\beta(1) = 1$, the ARMA representation (8.3) has a unit root, so that $u_t$ has no finite unconditional variance; this situation is called integrated GARCH (IGARCH). In the IGARCH(1, 1) case:

$$E[h_{T+H}] = h_T + \alpha_0 H,$$

which grows linearly with the forecast horizon $H$. However, the $u_t$ process is strictly stationary (provided $\alpha_0 > 0$, see Nelson, 1990).

The GARCH model can also be estimated with standardized Student-$t$ distributed errors (see §9.5). As the degrees of freedom go to infinity, this distribution approaches the standard normal.

We also consider the exponential GARCH, or EGARCH($p,q$), specification (see Nelson, 1991):

$$\log h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \{ \vartheta_1 \varepsilon_{t-i} + \vartheta_2 (|\varepsilon_{t-i}| - E|\varepsilon_t|) \} + \sum_{i=1}^p \beta_i \log h_{t-i}, \quad (8.4)$$

with $\alpha_1 = 1$. The log-formulation implies that $h_t$ is always nonnegative, regardless of the (possibly negative) parameter values; stationarity is now related to the roots of $1 - \beta(L)$, in particular to the question whether $\beta(1) < 1$ (but the implementation in PcGive does not impose restrictions on the EGARCH parameter space). When $\vartheta_1 \neq 0$, the model allows for an asymmetric response. In the normal EGARCH model, the errors follow a standard normal distribution. PcGive also allows the EGARCH model with a generalized error distribution (GED), which is defined as:

$$f(z) = \nu \exp \left( -\frac{1}{2} \frac{|z/\lambda|^\nu}{\nu} \right) \lambda^{2(1+1/\nu)} \Gamma(1/\nu), \quad -\infty < z < \infty \quad \nu > 0, \quad (8.5)$$

where $\Gamma(\cdot)$ is the gamma function, and $\lambda^2 = 2^{-2/\nu} \Gamma(1/\nu) / \Gamma(3/\nu)$. It contains the standard normal distribution as a special case, in which case $E|\varepsilon_t| = (2/\pi)^{1/2}$. When $\nu < 2$ the distribution has a thicker tail than the standard normal. PcGive estimates $\nu^* = \log \nu / 2$, which is zero for the standard normal distribution; in that case $\nu^* < 0$ implies thicker tails.

Both the GARCH and EGARCH models can be estimated with the conditional variance in the mean, with a choice between $h_t$, $h_t^{1/2}$ and $\log h_t$ entering the vector of regressors $x_t$.

\(^1\) Formally, PcGive imposes $\alpha(1) + \beta(1) < 1$ with strict inequality. However, numerically, the result can get so close as to being virtually indistinguishable from one. In that case the sum is assumed to be on the IGARCH boundary. By default, $\alpha(1) + \beta(1) > 1$ is excluded, although this restriction can be relaxed.
Finally, asymmetric and threshold versions of the GARCH model are available, where $h_t$ is specified as:

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i (u_{t-i} - \kappa_1)^2 + \kappa_2 D_{t-1} (u_{t-1} - \kappa_1)^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \quad (8.6)$$

where $D_{t-1} = 1$ if $u_{t-1} < \kappa_1$ and zero otherwise; $\kappa_1$ is the asymmetry parameter, and $\kappa_2$ the threshold parameter. Either or both can be zero: a pure threshold model has $\kappa_1 = 0$ (Glosten, Jagannathan, and Runkle, 1993), and a pure asymmetric GARCH model has $\kappa_2 = 0$.\(^2\)

In summary, the following models can be estimated:

GARCH
\[
\begin{align*}
\text{normal or } t \text{ errors} \\
\text{conditional variance in mean} \\
\text{threshold} \\
\text{asymmetric} \\
\text{regressors in conditional variance}
\end{align*}
\]

EGARCH
\[
\begin{align*}
\text{normal or GED errors} \\
\text{conditional variance in mean} \\
\text{regressors in conditional variance.}
\end{align*}
\]

---

\(^2\) The terminology we adopt here is somewhat arbitrary: both models actually allow for asymmetry.
Chapter 9

Tutorial on GARCH Modelling

9.1 Estimating a GARCH(1, 1) model

To illustrate the estimation of a GARCH(1, 1) model, we use the Deutschmark/British Pound exchange rate data from Bollerslev and Ghysels (1996). The data are daily percentage nominal returns, from 3-Jan-1984 to 31-Dec-1991, containing 1974 observations, and constructed from US dollar rates published in International Financial Statistics (by the International Monetary Fund).

The data are provided with PcGive as exch_debp.in7/exch_debp.bn7, so load this data set in OxMetrics. The variables in the data set are:

**DLxrDEBP** nominal daily returns on DM/UK£ exchange rate;

**FirstDay** a dummy variable indicating whether the previous day was a non-trading day (one for Mondays and post-holidays, and zero otherwise);

**sqrDLxrDEBP** the squared nominal daily returns.

In addition, there is the first variable called **Date**, which defines the calendar dates for the observations.1

Figure 9.1a displays the actual data, showing the volatility clustering which motivated the development of ARCH-type models. Whereas the sample autocorrelations of the daily returns are close to zero, see Figure 9.1b, the correlogram of the squared returns shows quite a different picture in Figure 9.1c: the squared returns seem significantly autocorrelated. This feature is typical for financial data. In this case there also appears to be some ‘seasonality’: at five and ten periods the autocorrelation is higher, expected to correspond to the 5-day working week cycle. This feature is the primary focus of Bollerslev and Ghysels (1996), but less so here.

---

1 The original data is undated, but we inferred dates from the calendar for that period. This may not be quite correct.
To access the GARCH estimation, activate PcGive (i.e. start the Model from Ox-Metrics), and select Models for financial data Garch Models using PcGive:

We assume that you already have some experience with using PcGive, in which case the remainder should look quite familiar. Press Formulate, and formulate the model as $DLxrDEBP$ on a Constant:
9.1 Estimating a GARCH(1, 1) model

This specifies the mean equation. Click on OK, and accept the default of a GARCH(1, 1) model:

OK takes us to the model estimation dialog, where the sample period can be adjusted:

accept the default again, which will start the estimation procedure. Despite the large sample size \( T = 1974 \), convergence is rapidly achieved:

VOL(1) Modelling DLxrDEBP by restricted GARCH(1,1)

   The dataset is: .\OxMetrics7\data\exch_debp.in7
   The estimation sample is: 1984-01-04 - 1992-01-02

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>robust-SE</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant X</td>
<td>-0.00619063</td>
<td>0.008376</td>
<td>0.008731</td>
<td>-0.709</td>
</tr>
<tr>
<td>alpha_0 H</td>
<td>0.0107597</td>
<td>0.001929</td>
<td>0.003123</td>
<td>3.44</td>
</tr>
<tr>
<td>alpha_1 H</td>
<td>0.153122</td>
<td>0.01940</td>
<td>0.02732</td>
<td>5.60</td>
</tr>
<tr>
<td>beta_1 H</td>
<td>0.805991</td>
<td>0.02184</td>
<td>0.03015</td>
<td>26.7</td>
</tr>
</tbody>
</table>

log-likelihood  -1106.60788 HMSE 5.52006
mean(h_t)       0.23018 var(h_t) 0.0408813
Chapter 9 Tutorial on GARCH Modelling

no. of observations  1974  no. of parameters  4
AIC.T     2221.21576  AIC    1.12523595
mean(DLxrDEBP)  -0.0164268  var(DLxrDEBP)  0.221018
alpha(1)+beta(1)  0.959113  alpha_i+beta_i>=0, alpha(1)+beta(1)<1

Initial terms of alpha(L)/[1-beta(L)]:

0.15312  0.12341  0.099471  0.080173  0.064619  0.052082
0.041978  0.033834  0.027270  0.021979  0.017715  0.014278

Used sample mean of squared residuals to start recursion
RobustSE based on analytical Information matrix and analytical OPG matrix
BFGS using analytical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence
Used starting values:

-0.016427  0.011051  0.78959  0.16041

The estimated model is a special case of (8.2):

\[ y_t = \mu + u_t, \]
\[ h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1}. \]

The output lists the coefficients and their standard errors. By default, robust standard errors are also reported, in which case those are used for the t-values as well. This is followed by the mean and variance of the conditional variance \( h_t \).

Next in the output is the log-likelihood, and the heteroscedasticity-adjusted mean square error (HMSE). More formally, the output comprises:

- Log-likelihood \( \hat{\ell} \) of the estimated model.
- The HMSE is defined as:

\[ \text{HMSE} = \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\hat{u}_t^2}{\hat{h}_t} - 1 \right)^2 = \frac{1}{T} \sum_{t=1}^{T} \left( \hat{\varepsilon}_t^2 - 1 \right)^2. \]

This is a loss function used by Bollerslev and Ghysels (1996) for model comparison.

- Mean and variance of the estimated conditional variance:

\[ \hat{h}_t = \frac{1}{T} \sum_{t=1}^{T} \hat{h}_t, \quad \text{var}(h_t) = \frac{1}{T} \sum_{t=1}^{T} \left( \hat{h}_t - \bar{h} \right)^2. \]

- AIC:

\[ \text{AIC.T} = -2\hat{\ell} + 2s, \quad \text{AIC} = -2\hat{\ell}/T + 2s/T, \]

where \( s \) the number of estimated parameters.

The last line of this block lists the value of \( \alpha(1) + \beta(1) \) under alpha(1)+beta(1). When this equals one, the model is IGARCH. In the default parameter space, the sum is constrained to be less than unity, thus ruling out IGARCH (this is indicated by the remainder of the line: \( \alpha_i + \beta_i \geq 0, \alpha(1) + \beta(1) < 1 \)).

Additional information on the dynamic properties is given when \( p > 1 \), in the form of \( 1 - \beta(1) \) and the inverse roots of \( 1 - \beta(L) \) (less then unity in absolute value indicates an invertible lag polynomial). This is followed by the 12 initial terms of
9.2 Evaluating the GARCH\((1, 1)\) model

The entries in the Test menu should again be familiar if you have used other PcGive packages before. From the Test menu select Graphic analysis. Because the mean equation only has an intercept, the graph of actual and fitted is not so interesting. Instead mark Conditional standard deviation, Residual autocorrelations, and Squared-residual autocorrelations, setting the length of the correlograms to 20:

![Graphic Analysis - GARCH Models]

the resulting graphs are in Figure 9.2. The first shows the scaled residuals \(\hat{\varepsilon}_t = \hat{u}_t/\hat{h}_t^{1/2}\), the next has \(\hat{h}_t^{1/2}\). The final two graphs show the autocorrelations of \(\hat{\varepsilon}_t\) and \(\hat{\varepsilon}_t^2\), with the latter showing that the autocorrelation in the squared residuals has indeed disappeared.

The Test Summary indicates serious non-normality (using the test of Doornik and Hansen, 1994), which is not surprising considering the many large residuals:

Descriptive statistics for scaled residuals:
- Normality test: \(\chi^2(2) = 435.66 [0.0000]\)**
- ARCH 1-2 test: \(F(2,1966) = 1.3066 [0.2710]\)
- Portmanteau(44): \(\chi^2(44) = 58.178 [0.0744]\)

As before, these tests are using the scaled residuals \(\hat{\varepsilon}_t\). The Test/Test menu allows for more detailed output:
Chapter 9 Tutorial on GARCH Modelling

Figure 9.2 GARCH(1, 1) model for daily returns on DM/UK£ exchange rate, residual ACF and ACF of squared residuals

A selective copy of the listing is:

Portmanteau(20): \( \chi^2(20) = 19.278 \ [0.5038] \)

Normality test for scaled residuals
Observations 1974
Mean -0.017759
Std.Devn. 0.99874
Skewness -0.34709
Excess Kurtosis 3.5220

The portmanteau statistics of \( \hat{\epsilon}_t \) and \( \hat{\epsilon}_t^2 - 1 \) are listed as a \( \chi^2 \) distribution with \( s - m \) degrees of freedom, where \( s \) is the lag length of the correlogram. For the scaled residuals \( m \) is the number of lags of the dependent variable, while for the squared scaled...
residuals: \( m = p + q \). This is approximate at best.

![Figure 9.3 Estimated density of scaled residuals of GARCH(1, 1) model](image)

Note that the scaled residuals \( \hat{\varepsilon}_t \) do not exactly have a zero mean, unlike in normal regression with a constant term (the Portmanteau statistic, however, uses the scaled residuals in deviation from their mean). The excess kurtosis of the residuals is also clearly visible in the residual density plot, see Figure 9.3. For comparison, a normal distribution is drawn with the same mean and variance, which has much thinner tails.

### 9.3 Recursive estimation of the GARCH(1, 1) model

Recursive estimation is a useful tool in assessing parameter constancy. PcGive allows any GARCH model to be estimated recursively, by clicking on Recursive estimation in the Estimation model dialog. It is advisable to set the number of initializations not too small: we choose 200 here. Recursive estimation starts from the full sample estimates, reducing the sample size one at a time. By using the parameter estimates at the previous step, recursive estimation is surprisingly quick: the current procedure takes only a couple of minutes, although it involves nearly 1800 GARCH model estimations.

When the recursive estimation is finished, use Test/Recursive Graphics to graph the results:

![Recursive Graphics](image)

Choose Coefficients for all coefficient and their standard errors, and alpha(1)+beta(1) for \( \hat{\alpha}(1) + \hat{\beta}(1) \) (which is just \( \hat{\alpha}_1 + \hat{\beta}_1 \) in the GARCH(1, 1) model).
Note from Figure 9.4a that early on the confidence interval includes some negative values, despite the fact that we impose $\alpha_0 > 0$. This is a consequence of using an asymptotic normal approximation. The estimation is in terms of $\psi_0$ with $\alpha_0 = \exp(\psi_0)$ which is always positive, and then transforming back afterwards. A more appropriate confidence interval would be based on $[\exp(\hat{\psi}_0 - 2s_{\hat{\psi}}), \exp(\hat{\psi}_0 + 2s_{\hat{\psi}})]$, which will not include points outside the parameter space. We leave this to a future version.

We can re-estimate without restrictions (but $\alpha_0 > 0$ is always imposed), by selecting Unrestricted estimation in the Model Settings stage:
9.3 Recursive estimation of the GARCH(1, 1) model

Figure 9.5 compares the two parametrizations. The fact that, when moving backward, the \( \hat{\alpha}_1 + \hat{\beta}_1 = 1 \) ‘period’ lasts longer in the first graph than \( \hat{\alpha}_1 + \hat{\beta}_1 > 1 \) in the second may be an artefact from the recursive estimation: with starting values close to IGARCH, the model has a tendency to converge to IGARCH (although the recursive procedure tries to robustify against this).

![Graph showing recursively estimated GARCH(1, 1) model with and without \( \alpha_1 + \beta_1 < 1 \)]
Bollerslev and Ghysels (1996) consider a model where the trading-day dummy variable enters the variance equation:

\[ h_t = \alpha_0 + \alpha_0^*d_t + \alpha_1u_{t-1}^2 + \beta_1h_{t-1}. \]

The trading day variable has value one for a first trading day, in which case the variance intercept is \( \alpha_0 + \alpha_0^* \), and a zero otherwise, with intercept \( \alpha_0 \).

To estimate such a model, add \textit{FirstTrade} to the model, clear the status, and then mark it as H: X in \( h_t \) as shown in the capture:

\begin{center}
\textbf{Data Selection - GARCH Models - exch_debp.in7}
\end{center}

In the Model Settings reset the GARCH parameter restrictions to Impose stationarity and \( \alpha+\beta \geq 0 \):

\begin{itemize}
  \item VOL( 4) Modelling DLxrDEBP by restricted GARCH(1,1)
  \item The dataset is: .\OxMetrics7\data\exch_debp.in7
  \item The estimation sample is: 1984-01-04 - 1992-01-02
\end{itemize}

\begin{verbatim}
<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std_ERROR</th>
<th>robust-SE</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant X</td>
<td>-0.006723 0.008109 0.008579 -0.784 0.433</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FirstTrade H</td>
<td>0.056062 0.009863 0.02260 2.48 0.013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alpha_0 H</td>
<td>1.39937e-011</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alpha_1 H</td>
<td>0.180019 0.02402 0.04233 4.25 0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beta_1 H</td>
<td>0.773468 0.02535 0.05253 14.7 0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log-likelihood</td>
<td>-1090.21648</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean(h_t)</td>
<td>0.23289 var(h_t) 0.0459151</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no. of observations</td>
<td>1974 no. of parameters 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC.T 2190.43296 AIC 1.10964183</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alpha(1)+beta(1) 0.953486 alpha_i+beta_i&gt;=0, alpha(1)+beta(1)&lt;1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
\end{verbatim}

The \( \hat{\alpha}_0 \) coefficient is very close to zero, indicating that it is very close the boundary of the parameter space, which always imposes \( \alpha_0 > 0 \). The standard error may therefore not be meaningful (at can also happen that the standard error is omitted in the output in that case). Here \( \hat{\alpha}_0 \) is effectively zero, which is the value for any day which is not a first trading day. If so desired, this restriction can be avoided by changing the definition of the dummy variable, for example:
NextTrade = 1 - FirstTrade;

Re-estimating with this variable gives an intercept of 0.059 on opening days, and −0.003 afterwards:

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>robust-SE</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant X</td>
<td>-0.00694288</td>
<td>0.008048</td>
<td>0.008651</td>
<td>-0.803</td>
</tr>
<tr>
<td>NextTrade H</td>
<td>-0.0614605</td>
<td>0.01099</td>
<td>0.02065</td>
<td>-2.98</td>
</tr>
<tr>
<td>alpha_0 H</td>
<td>0.0585031</td>
<td>0.009669</td>
<td>0.01996</td>
<td>2.93</td>
</tr>
<tr>
<td>alpha_1 H</td>
<td>0.169347</td>
<td>0.02469</td>
<td>0.04603</td>
<td>3.68</td>
</tr>
<tr>
<td>beta_1 H</td>
<td>0.792415</td>
<td>0.02914</td>
<td>0.06022</td>
<td>13.2</td>
</tr>
</tbody>
</table>

log-likelihood -1089.75782

HMSE 5.01853

alpha(1)+beta(1) 0.961766

alpha_i+beta_i>=0, alpha(1)+beta(1)<1

The impact of the addition of the trading-day variable can be illustrated by the effect it has on the forecast error bands. In the model with just an intercept, forecasts can be made without the need to adjust the database. However, when the trading day variable is used, the database must be extended to include trading-day observations for the forecast sample. We hypothesized the following dates at the sample end (indicating that the last observation is 2-Jan-1992):²

<table>
<thead>
<tr>
<th>FirstTrade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
</tr>
<tr>
<td>Tuesday</td>
</tr>
<tr>
<td>Wednesday</td>
</tr>
<tr>
<td>Thursday</td>
</tr>
<tr>
<td>Friday</td>
</tr>
<tr>
<td>Monday</td>
</tr>
<tr>
<td>Tuesday</td>
</tr>
<tr>
<td>Wednesday</td>
</tr>
<tr>
<td>Thursday</td>
</tr>
<tr>
<td>Friday</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>23-Dec</th>
<th>24-Dec</th>
<th>25-Dec</th>
<th>26-Dec</th>
<th>27-Dec</th>
<th>30-Dec</th>
<th>31-Dec</th>
<th>1-Jan</th>
<th>2-Jan</th>
<th>3-Jan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuesday</td>
<td>closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wednesday</td>
<td>closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thursday</td>
<td>closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friday</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monday</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuesday</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wednesday</td>
<td>closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thursday</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friday</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The last italic line is the first observation added to the data. After that we have added four times 1 0 0 0 0 to the data set.

These ‘future’ observations for the trading day variables have already been added to the dataset.

²The statusbar shows the day of a particular date if the database cursor is on an observation.
Use Test/Forecast with 21 forecasts and pre-sample observations:

![GARCH Model Forecasting Interface]

Figure 9.6 shows the effect of the dummy on the conditional variance (the forecast bands are $\pm 2\hat{h}_t^{1/2}$). The consequences of the zero intercept in $h_t$ is that the forecasted $\hat{h}_t$ will collapse to zero if continuous trading were introduced (looking at Figure 9.6b, it is clear that the positive intercept on the first trading day pushes the variance away from zero each time). A negative intercept would collapse to zero faster. This seems an undesirable feature of the current model.

![GARCH Model Forecast Comparison]

Figure 9.6  Forecasts from GARCH(1, 1) model with and without trading day effect
9.5 GARCH(1, 1) with Student t-distributed errors

PcGive allows the estimation of GARCH models with Student-t distributed errors, as proposed by Bollerslev (1987). This extends model (8.2), for the current specification:

\[ y_t = \mu + h^{1/2} \varepsilon_t, \quad \varepsilon_t | \mathcal{F}_{t-1} \sim t[\nu], \]
\[ h_t = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 h_{t-1}. \]

Here \( t[\nu] \) denotes the standardized t-distribution with \( \nu \) degrees of freedom, such that \( \mathbb{E}[\varepsilon_t^2 | \mathcal{F}_{t-1}] = 1 \), and \( \mathbb{E}[\varepsilon_t^4 | \mathcal{F}_{t-1}] = 3(\nu - 2)(\nu - 4)^{-1} \). The distribution tends to the standard normal when \( \nu \to \infty \).

Estimation in PcGive requires the selection of Non-normal error distribution in the Model Settings:

The estimated model is an IGARCH model, because \( \hat{\alpha}_1 + \hat{\beta}_1 = 1 \). The degrees of freedom are estimated at 4.25, which is quite far from normal:

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>robust-SE</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant X</td>
<td>0.00365860</td>
<td>0.006947</td>
<td>0.006960</td>
<td>0.526</td>
</tr>
<tr>
<td>NextTrade H</td>
<td>-0.0378227</td>
<td>0.01213</td>
<td>0.01436</td>
<td>-2.63</td>
</tr>
<tr>
<td>alpha_0 H</td>
<td>0.0328969</td>
<td>0.01022</td>
<td>0.01253</td>
<td>2.63</td>
</tr>
<tr>
<td>alpha_1 H</td>
<td>0.128697</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beta_1 H</td>
<td>0.871303</td>
<td>0.02547</td>
<td>0.03950</td>
<td>22.1</td>
</tr>
<tr>
<td>student-t df</td>
<td>4.25082</td>
<td>0.3350</td>
<td>0.3463</td>
<td>12.3</td>
</tr>
</tbody>
</table>

log-likelihood -984.014566 HMSE 5.57213
alpha(1)+beta(1) 1 alpha(1)+beta(1)>=0, alpha(1)+beta(1)<1

Figure 9.7 Forecasts from GARCH(1, 1) and GARCH-t(1, 1) models
The increase in the log-likelihood is large, strongly rejecting the previous GARCH model with normal errors. Because $\nu > 2$, the t-value for the degrees of freedom cannot be interpreted. Figure 9.7 shows the impact on the forecasts, with the dotted line corresponding to the most recent model with t[4.25]-distributed errors. The coverage probabilities are set at 90%, with normal critical value of 1.65, and 2.1 for t(4.25).

### 9.6 EGARCH(1, 1) GED-distributed errors

An EGARCH(1, 1) with errors which follow a generalized error distribution is estimated by clicking the EGARCH check box in Model Settings. Assuming that you follow on from the previous section, Non-normal error distribution will still be selected. The results are:

\[
\begin{array}{cccccc}
\text{Coefficient} & \text{Std. Error} & \text{robust-SE} & \text{t-value} & \text{t-prob} \\
\text{Constant X} & -0.000584468 & 0.006983 & 0.007296 & -0.0801 & 0.936 \\
\text{NextTrade H} & -0.302348 & 0.1011 & 0.1245 & -2.43 & 0.015 \\
\text{alpha_0 H} & 0.129214 & 0.07372 & 0.07433 & 1.74 & 0.082 \\
\text{eps[-1] H} & -0.0310867 & 0.02154 & 0.02209 & -1.41 & 0.160 \\
|\text{eps[-1]}| H & 0.332165 & 0.05099 & 0.07886 & 4.21 & 0.000 \\
\text{beta_1 H} & 0.941306 & 0.041743 & 0.02829 & 33.3 & 0.000 \\
\text{GED log(nu/2)} & -0.539225 & 0.04034 & 0.04462 & -12.1 & 0.000 \\
\end{array}
\]

\[
\begin{array}{cccc}
\text{log-likelihood} & -995.581391 & \text{HMSE} & 6.07972 \\
\text{mean(h_t)} & 0.229904 & \text{var(h_t)} & 0.0377776 \\
\text{no. of observations} & 1974 & \text{no. of parameters} & 7 \\
\text{AIC.T} & 2005.16278 & \text{AIC} & 1.01578662 \\
\text{mean(DLxrDEBP)} & -0.0164268 & \text{var(DLxrDEBP)} & 0.221018 \\
\end{array}
\]

Used sample mean of squared residuals to start recursion

Robust-SE based on numerical Hessian matrix and numerical OPG matrix

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005):

Strong convergence

Used starting values:

\[
\begin{array}{cccccc}
-0.016427 & 0.00000 & 1.1108 & 0.78959 & 0.00000 & 0.16041 \\
0.00000 & \\
\end{array}
\]

In terms of (8.4), $\text{eps[-1]}$ is $\vartheta_1$ and $|\text{eps[-1]}|$ is $\vartheta_2$. The GED(2) corresponds to the standard normal, and, since $\log(2/2) = 0$, that is again rejected in favour of a distribution with thicker tails.

---

3We use a specific to general approach here for illustrative purposes: notice how the opposite route led to us having to discard results as now rejected.
9.7 GARCH in mean

In the GARCH-M model, see Engle, Lilien, and Robins (1987), the conditional variance enters the mean equation:

\[ y_t = x_t' \zeta + \delta h_t^\lambda + u_t. \]

In PcGive \( \lambda = 1, \lambda = \frac{1}{2} \), or \( \log h_t \) can be used. We use an example along the lines of Engle, Lilien, and Robins (1987). From six-months secondary market rates on US Treasury bill rates \( (R_t, \text{called } tbsm6m \text{ in the database}) \) and three-month rates \( (r_t, tbsm3m) \) we construct the excess holding yield as:

\[ y_t = 100 \left[ \frac{(1 + R_t)^2}{1 + r_{t+1}} - (1 + r_t) \right]. \]

The data were downloaded from the US Federal Reserve, and are monthly from 1958(12) to 2000(12). The last observation for \( y_t \) is missing owing to the transformation, and the data are supplied as USyieldm.in7/USyieldm.bn7.

The GARCH-M(1,1) model to be formulated is:

\[
\begin{align*}
    y_t &= \zeta_0 + \zeta_1 y_{t-1} + \zeta_2 y_{t-3} + \zeta_3 h_t^{1/2} + u_t, \\
    h_t &= \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1}.
\end{align*}
\]

The constant and lagged dependent variables are added at the first stage of model formulation. Adding the \( h_t^{1/2} \) term requires expanding the GARCH variations group in Model Settings, and clicking on \( \sqrt{h_t} \):

The output reports that \( \hat{\alpha}_1 + \hat{\beta}_1 = 1 \). The coefficient of the conditional standard deviation in mean is \( \hat{\zeta}_3 = 0.228 \), and just significant. Figure 9.8 displays the graphical analysis of the estimated model, showing that the model has particular problems during the turbulence of the early eighties.

VOL( 8) Modelling y by restricted GARCHM(1,1,"sqrt")
The dataset is: ..\OxMetrics7\data\USyieldm.in7
The estimation sample is: 1959 (3) to 2000 (11)
The recursive estimates (with 200 observations for initialization) in Figure 9.9 reveal that the parameter estimates are constant, but also shows that the estimated model is IGARCH throughout (remember that this is a boundary of the parameter space in the default settings), which is causing the peculiar behaviour for the standard errors of $\alpha_1$. Of course, when $\alpha(1) + \beta(1) = 1$ is a boundary of the parameter space, the symmetric standard error band around unity does not make sense.
Unrestricted estimation has not much effect on the log-likelihood, and has $\hat{\alpha}_1 + \hat{\beta}_1 = 1.026$:

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>robust-SE</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$ Y</td>
<td>0.521145</td>
<td>0.04030</td>
<td>0.04353</td>
<td>12.0</td>
</tr>
<tr>
<td>$y_3$ Y</td>
<td>0.0863785</td>
<td>0.03967</td>
<td>0.04399</td>
<td>1.96</td>
</tr>
<tr>
<td>Constant X</td>
<td>0.0529767</td>
<td>0.01983</td>
<td>0.01567</td>
<td>3.38</td>
</tr>
<tr>
<td>alpha_0 H</td>
<td>0.00204980</td>
<td>0.001175</td>
<td>0.001562</td>
<td>1.31</td>
</tr>
<tr>
<td>alpha_1 H</td>
<td>0.284553</td>
<td>0.05733</td>
<td>0.07679</td>
<td>3.71</td>
</tr>
<tr>
<td>beta_1 H</td>
<td>0.741698</td>
<td>0.04189</td>
<td>0.05724</td>
<td>13.0</td>
</tr>
<tr>
<td>sqrt(h_t) X</td>
<td>0.211909</td>
<td>0.08568</td>
<td>0.07715</td>
<td>2.75</td>
</tr>
</tbody>
</table>

$\text{log-likelihood} = -141.971736$ $\text{HMSE} = 2.81231$
$\text{mean(h_t)} = 0.283496$ $\text{var(h_t)} = 0.366638$
$\text{alpha(1)+beta(1)} = 1.02625$

The recursive estimates from the unrestricted model in Figure 9.10 confirm that $\hat{\alpha}_1 + \hat{\beta}_1 > 1$ throughout the estimation sample. The other parameters are not much affected by the restriction. The impact of this relaxation of the parameter space is most noticeable on forecasts. Figure 9.11a gives the two-year forecasts with standard error bands from the IGARCH-M model, followed by the forecasts of the unconditional variance. The next two graphs repeat this for the unconstrained model, which has $\hat{\alpha}_1 + \hat{\beta}_1 = 1.026$. The latter show rapid growth of the forecasts of $h_t$, which, in these models, feeds back into the forecasts of $y_t$. Both models seem quite unsatisfactory.
Finally, to illustrate the possibilities, we add a threshold and asymmetry effect. When estimating this model, the in-mean effect becomes insignificant. Deleting this from the model results in the following specification:

\[
\begin{align*}
y_t &= \zeta_0 + \zeta_1 y_{t-1} + \zeta_2 y_{t-3} + u_t, \\
h_t &= \alpha_0 + \alpha_1 (u_{t-1} - \kappa_1)^2 + \beta_1 h_{t-1} + \kappa_2 u_{t-1} I(u_{t-1} < \kappa_1),
\end{align*}
\]

where \( I(\cdot) \) is the indicator function. Estimation (keeping unrestricted estimation as set in the previous section) yields:

\text{VOL(9) Modelling } y \text{ by unrestricted ATGARCH(1,1) (usyieldm.in7)}

The estimation sample is: 1959 (3) to 2000 (11)
The $\kappa_1$ parameter is listed as asymmetry, and $\kappa_2$ as threshold. Recursive estimation gives a few warning messages of the kind ‘invertgen: singular matrix’, which indicates that there was a problem evaluating the variance-covariance matrix of the pa-
parameter estimates. This is causing the blips in the error bands which mainly occur in the early eighties in the recursive graphs of Figure 9.12. The forecasts in Figure 9.13 are quite different to those from the GARCH-M model.

![Figure 9.13 Forecasts for TAGARCH(1, 1) model](image)

### 9.9 GARCH batch usage

As with most OxMetrics modules, batch code is automatically generated in the background. Activating the batch editor, shows the most recently estimated model, as displayed on the previous page.

The Load history button will show the batch code for all estimated models. The
Batch commands that are specific to GARCH estimation are listed in §10.3. See volume3_garch.fl in the batch folder for an example. This batch program estimates the models from this chapter, with the recursive ones commented out.

Starting from OxMetrics 5, Ox code can also be generated. Provided Ox Professional is available, this code can be run, either from OxMetrics, or from OxEdit or the command line.

The volume3_garch.ox code in the batch folder provides an example. The only changes that were made to the generated code consisted of making the paths to the data files relative and commenting out the recursively estimated runs.
Chapter 10

GARCH Implementation Details

10.1 GARCH model settings

The default is a GARCH(1, 1) model with normal errors, see (8.2). If desired, other values for \( p \) and \( q \) can be selected.

The Model Settings dialog also has a range of options (the default is marked below with a *):

1. **Main options**
   - (a) Order \( p \) and \( q \)
   - (b) choice between EGARCH\((p, q)\), see (8.4), or GARCH\((p, q)\), see (8.2)
   - (c) Non-normal error distribution
     This is either a GARCH model with student-t distributed errors or an EGARCH model with errors which have a generalized error distribution.

2. **GARCH variations**
   - (a) Asymmetric GARCH, see (8.6)
   - (b) Threshold GARCH, see (8.6)
     Selecting both results in a threshold asymmetric GARCH model,
   - (c) Conditional variance in mean
     - No conditional variance in mean
     - \( h_t \) in mean
     - \( \sqrt{h_t} \) in mean
     - \( \log(h_t) \) in mean

3. **GARCH parameter restrictions**
   - (a) Unrestricted estimation
     This imposes no restrictions, and could in rare cases lead to negative variance estimates.
   - (b) Impose conditional variance \( \geq 0 \)
     Nelson and Cao (1992) argued that imposing all coefficients to be nonnegative is overly restrictive, and that negative estimates occur in practice (they list several examples). Subsequently, He and Teräsvirta (1999) have shown that such nega-
tive coefficients allow for richer shapes of the autocorrelation function. Doornik and Ooms (2008) discuss the implementation of these restrictions.

(c) *Impose stationarity and alpha+beta $\geq 0$

Doornik and Ooms (2008) also considered restrictions in terms of (8.3). Defining $m = \max(p, q)$, $\beta_i = 0$ for $i > p$, $\alpha_i = 0$ for $i > q$:

$$\begin{align*}
\alpha_0 &\geq 0, \\
\alpha_i + \beta_i &\geq 0, \quad \text{for } i = 1, \ldots, m, \\
\sum_{i=1}^{m} \alpha_i + \beta_i &< 1. 
\end{align*}$$

In terms of (8.3), these restrictions imply that the unconditional variance exists, and is always positive. Formally, PcGive imposes $\alpha(1) + \beta(1) < 1$ with strict inequality. However, numerically, the result can get so close as to being virtually indistinguishable from one. In that case the sum is assumed to be on the IGARCH boundary.

(d) Impose alpha,beta $\geq 0$

Bollerslev (1986) proposed the GARCH model with $\alpha_0 \geq 0$, $\alpha_i \geq 0$, and $\beta_i \geq 0$. This ensures that $h_t \geq 0$.

4. **Start-up of GARCH variance recursion.**

Evaluation of the likelihood requires pre-sample values for $u_t^2$ and $h_t$. PcGive implements two methods:

(a) *Use mean variance

Bollerslev (1986) suggested using the mean of the squared residuals:

$$u_i^2 = h_i = T^{-1} \sum_{t=1}^{T} u_t^2, \quad \text{for } i \leq 0.$$  

(10.2)

In EGARCH, the recursion is started with the logarithm of the mean of the squared residuals.

(b) Estimate missing variances

This amounts to adding the missing $u_1^2 \ldots u_m^2$ as parameters which are to be estimated. This make the likelihood derivatives more complex, and PcGive will only use numerical derivatives if this method is selected (see below).

5. **Preferred covariance estimator**

(a) Second derivatives

This uses minus the inverse of the numerical second derivative of the log-likelihood function.

(b) *Information matrix

The information matrix is computed analytically, but only for standard GARCH models.

(c) Outer product of gradients

In addition, the robust standard errors are printed by default when the information matrix $J$ is available. These are of the form suggested by Bollerslev and Wooldridge (1992): $\hat{J}^{-1} \hat{G} \hat{J}^{-1}$, where $\hat{G}$ is the outer product of the gradients.
A comparison of various estimators is given in Fiorentini, Calzolari, and Panattoni, 1996.

6. Search for global maximum after initial estimation
Especially when \( p > 1 \), it is possible that the likelihood has multiple optima. This final set of advanced options allows for a search from random starting values. Because each of these involves maximization of the likelihood, this option can be time consuming. See Doornik and Ooms (2008) for an application.

10.2 Some implementation details

1. Maximization technique
PcGive only offers BFGS (see e.g. Fletcher, 1987 or Gill, Murray, and Wright, 1981), because it is our preferred method for numerical optimization. BFGS avoids the need for second derivatives, while being one of the most robust methods available. This is supplemented by a line search when, at an iteration step, the likelihood does not increase. BFGS was not considered by Fiorentini, Calzolari, and Panattoni (1996), but we found 100% convergence when replicating their Table 1 with 1000 replications (requiring about 17 iterations on average, whether starting from the DGP values, or from a starting value routine).

2. Starting values for the parameters
PcGive uses the method of Galbraith and Zinde-Walsh (1997) applied to the squared data (after removing regressors in the mean). If necessary, the resulting parameter values are reduced to ensure that the unconditional variance exists.

3. Use of analytical derivatives
Numerical derivatives are more convenient, but less accurate than analytical derivatives (see Fiorentini, Calzolari, and Panattoni, 1996). The latter are to be preferred, but convenience often dictates the use of the former. In simple GARCH models, we found numerical derivatives sufficiently effective, with model estimation taking the same amount of time, and convergence achieved as frequently.
PcGive uses analytical derivatives, except (1) when the Hessian matrix is required for the variance-covariance matrix, (2) when the initial values of the conditional variances are added as parameters, (3) for EGARCH-type models and (4) for conditional variance in mean models.
10.3 GARCH batch commands

\[ T[A]GARCH[M][GARCH](p, q); \]

\[ EGARCH[M][GED](p, q); \]

Specifies the model to estimate, with the terms in [ ] optional. Some examples are:
\[ GARCH(1,1), \ TAGARCH(1,1), \ GARCHM(1,1), \ EGARCH_GED(1,1). \] For in mean,
a third argument can be "sqrt", or "log", the default being \( h_t \) when no third
argument is present.

\[ \text{forecast}(ctforc); \]
prints \( ctforc \) forecasts.

\[ \text{estimate}("ML", year1=-1, period1=0, year2=-1, period2=0, init= 0); \]
Estimate a system. The presence of default arguments implies that the shortest ver-
sion is just: \( \text{estimate()} \), which estimates by OLS using the maximum possible
sample, and no forecasts. Similarly, a call to \( \text{estimate}("ML", 1950, 1) \) corre-
sponds to \( \text{estimate}("ML", 1950, 1, -1, 0, 0, 0) \).
\[ year1(period1) – year2(period2) \] is the estimation sample. Setting year1 to \(-1\) will
result in the earliest possible \( year1(period1) \), setting year2 to \(-1\) will result in the
latest possible \( year2(period2) \).
\[ init \] is the number of observations to use for initialization of recursive estimation;
no recursive estimation is performed if \( init = 0 \).

If the database is dated, specify the start date in the \( year1 \) argument and end
date in \( year2 \), with both periods set to zero. An example is \( \text{estimate}("ML", 1984-01-04, 0, 1992-01-02, 0) \).

\[ \text{module}("PcGive"); \]
Starts the PcGive module. If PcGive is already running, this batch command is not
required.

\[ \text{option}"option", argument"; \]

<table>
<thead>
<tr>
<th>option</th>
<th>argument</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>maxit</td>
<td>maximum number of iterations</td>
<td>default: 1000,</td>
</tr>
<tr>
<td>print</td>
<td>print every # iteration</td>
<td>0: do not print,</td>
</tr>
<tr>
<td>compact</td>
<td>compact or extended output</td>
<td>0 for off, 1 on,</td>
</tr>
<tr>
<td>strong</td>
<td>strong convergence tolerance</td>
<td>default: 0.005,</td>
</tr>
<tr>
<td>weak</td>
<td>set weak convergence tolerance</td>
<td>default: 0.0001,</td>
</tr>
</tbody>
</table>

\[ \text{package}("GARCH"); \]
Use this command to select the GARCH package.

\[ \text{progress}; \]
Reports the modelling progress.

\[ \text{store}"name", "rename"=""; \]
Use this command to store residuals, etc. into the database, the default name is
used. Note that if the variable already exists, it is overwritten without warning. The
name must be one of:
- residuals  unscaled residuals
- scaledres scaled residuals
- condvar conditional variance
- forecasts forecasted values (after using forecast command)
- forcsse forecast standard errors

The optional second argument replaces the default name. For example, `store("residuals")` stores the residuals under the name Residual; `store("residuals", "xyz")` stores them under the name xyz.

system { Y=...; X=...; H=...; }

Specify the system, consisting of the following components:
- Y endogenous variable;
- X non-modelled variables;
- H regressor in conditional variance.

The variables listed are separated by commas, their base names (that is, name excluding lag length) must be in the database. If the variable names are not a valid token, the name must be enclosed in double quotes.

The following special variables are recognized: Constant, Trend, Seasonal and CSeasonal.

test("test", lag1=0, lag2=0);

Performs a specific test using the specified lag lengths.

```
"ar" test for autocorrelated errors from lag1 to lag2;
"arch" ARCH test up to order lag1;
"normal" normality test;
"rescor" residual correlogram up to lag lag1;
"rescorsqr" residual correlogram of squares up to lag lag1;
```

testlinres {...}

Test for linear restrictions. The content is the matrix dimensions followed by the \((R : r)\) matrix.

testsummary;

Do the test summary.
Part V

Time Series Models (ARFIMA)

with Marius Ooms
Chapter 11

Introduction to Time Series Models

The fractionally-integrated autoregressive-moving average model, denoted ARFIMA\((p,d,q)\) can be used for the statistical analysis of a univariate time series \(y_t\) with long memory. The more familiar ARMA (or ‘Box–Jenkins’) models (when \(d = 0\)) or ARIMA (when \(d\) is a positive integer) are special cases. The ARFIMA\((p,d,q)\) model allows one to model the long-run behaviour of a time series in a flexible way. Empirical modelling involves identification, estimation and testing. In the identification stage of ARIMA-modelling one determines the integer part of the order of differencing, \(d\), and the orders of the autoregressive and the moving-average parts of the model, \(p\), and \(q\). In the ARFIMA-model one can estimate \(d\) and compute a confidence interval. The ARFIMA model allows for likelihood ratio and Wald tests for the null of short-memory stationarity \((d = 0)\), as well as for the null of unit-root non-stationarity \((d = 1)\) against long memory and intermediate memory alternatives. These tests are easy to apply since they have conventional chi-squared limit distributions.

The tests complement widely applied Dickey–Fuller type tests for \(d = 1\) against short memory alternatives and tests for \(d = 0\) against \(d = 1\) based on variances of cumulative sums. PcGive’s parameter instability test applied to the constant term of a regression model can be interpreted as a test for \(d = 0\) with fixed mean, against \(d = 1\) with a random walk mean.

From the early nineteen sixties onwards, when Mandelbrot suggested long-memory models for economic time series, there has been a steady growth in the literature on the subject. Robinson (1994) and Baillie (1996) provided useful surveys of the developments in the econometric modelling of long memory; Beran’s monograph, Beran (1994), discusses most of the central issues, including forecasting. The implementation in PcGive is derived from the Arfima package for Ox by Doornik and Ooms (1999), while an application and a more extensive discussion can be found in Doornik and Ooms (2004).

We write the ARFIMA\((p,d,q)\) model as:

\[
\Phi (L) (1 - L)^d (y_t - \mu_t) = \Theta (L) \varepsilon_t, \quad t = 1, \ldots, T.
\]

101
where $\Phi(L) = (1 - \phi_1 L - \ldots - \phi_p L^p)$ is the autoregressive polynomial and $\Theta(L) = (1 + \theta_1 L + \ldots + \theta_q L^q)$ is the moving average polynomial in the lag operator $L$; $p$ and $q$ are integers, $d$ is real. For example, the ARFIMA$(1, d, 1)$ model is:

$$(1 - L)^d z_t = \phi_1 (1 - L)^d z_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1},$$

where

$z_t = y_t - \mu_t$.

The notation in textbooks often differs in terms of the sign of the parameters: we follow the regression-model notation.

$(1 - L)^d = \Delta^d$ is the fractional difference operator defined by the following binomial expansion:

$$(1 - L)^d = \sum_{j=0}^{\infty} \delta_j L^j = \sum_{j=0}^{\infty} \binom{d}{j} (-L)^j.$$

We assume $\varepsilon_t \sim \text{NID}[0, \sigma_\varepsilon^2]$, and write $\mu_t$ for the mean of $y_t$. The ARMA-part of the model is assumed invertible and stationary: all roots of $\Theta(z) = 0$ and $\Phi(z) = 0$ are outside the unit circle. In addition, $\Theta(z) = 0$ and $\Phi(z) = 0$ do not have common roots. We say that $z_t = y_t - \mu_t$ is $I(d)$, integrated of order $d$. The zero mean ARFIMA$(p, d, q)$ process $z_t$ is covariance stationary if $d < 0.5$, see Hosking (1981), with autocovariance function that decays hyperbolically. The process is long memory in the case $0 < d < 0.5$. For $-0.5 < d < 0$ the process is called intermediate memory or 'overdifferenced', see Brockwell and Davis (1993) and e.g. Chang and Dickey (1994). We assume $d > -1$, which makes the process $z_t$ invertible, see Odaki (1993).

We provide three estimation methods, Exact Maximum Likelihood (EML), Modified Profile Likelihood (MPL) and nonlinear least squares (NLS). By definition, EML and MPL impose $-1 < d < 0.5$. MPL is preferred over EML if the model includes regressor variables and the sample is not very large. NLS allows for $d > -0.5$ and can be used to estimate models for non-stationary processes directly, without a priori differencing. NLS estimation is usually fast.

The next chapter shows how PcGive can be used to estimate ARFIMA models. We discuss implementation details in Chapter 13.
Chapter 12

Tutorial on ARFIMA Modelling

The data set for this tutorial is \texttt{UKrpiq.in7}/\texttt{UKrpiq.bn7} consisting of the UK retail price index (all items). Inflation is constructed as the growth rate of the price index:

\[ i^q_t = 400 \Delta \log(P_t) \] quarterly annualized percentage,
\[ i^a_t = 100 \Delta_4 \log(P_t) \] year to year percentage change.

In OxMetrics Algebra terms:

\begin{align*}
\text{LRPI} &= \log(\text{RPI});
\text{InflatQ} &= 400 \ast \text{diff}(\text{LRPI}, 1);
\text{InflatA} &= 100 \ast \text{diff}(\text{LRPI}, 4);
\end{align*}

The data are quarterly and run from 1948-3 and 1949-1 to 2000-4 after differencing. The data source is the UK Office of National Statistics, check \url{www.statistics.gov.uk} to obtain the latest observations. The variable of interest is \textit{InflatQ}, which is quarterly inflation at annual rates based on the RPI index. Figure 12.1 shows quarterly inflation and annual inflation, both at annual rates. Note how annual inflation lags annualized quarterly inflation.

Our first objective here is to show how an \textit{ARFIMA} model is estimated in PcGive. The second objective is to forecast inflation using \textit{ARFIMA} models. Therefore, we immediately estimate a model which is similar to that suggested by Doornik and Ooms (2004), without going through the normal exploratory data analysis and early modelling stages. The model is a seasonal \textit{ARFIMA} model:

\begin{align*}
(1 - L)^d z_t &= \phi_4 (1 - L)^d z_{t-4} + \varepsilon_t + \theta_4 \varepsilon_{t-4}, \\
y_t &= \beta_0 + \beta_1 Q_1 + \beta_2 Q_2 + \beta_3 Q_3 + \beta_4 D_{75}Q_2 + \beta_5 D_{79}Q_3 + z_t.
\end{align*}

\texttt{Q1}, \texttt{Q2}, and \texttt{Q3} are seasonal dummy variables, with, for \texttt{Q1}, a one in the first quarter, and zero otherwise. The remaining seasonals are defined accordingly.

The two dummy variables capture two peaks in inflation: at 1975Q2, when annualized quarter-to-quarter inflation was 35\%, and 1979Q3, when it was over 25\%. Our
model will not be able to track these extreme values, so we allow for a different mean\( \mu_t \) for these observations. Without these dummies residual-based tests signal strong non-normality.

When using the dummy variables, we need to extend the database to include observations (just zeros) over the forecast period. The extended data set runs to 2002Q4, allowing for 8 out-of-sample forecasts.

We now assume that you have loaded the UKrpiq.in7 data set into OxMetrics. Then start the Model dialog and first select Models for time-series data and then the model class ARFIMA Models using PcGive:

![PcGive - Models for time-series data](image)

We expect that you already have some experience with using PcGive, in which case the remainder should look quite familiar. Select the Formulate button in the dialog and formulate the model as \( \text{InflatQ} \) on a \textit{Constant}, the three seasonals (the three are...
automatically added when clicking on Seasonal) and the two dummy variables:

This specifies the mean equation: $\mu_t$ in (11.1). Click on OK, to see:

Starting from below, the first issue is the treatment of the mean. Here we already have a Constant as regressor in the mean, and therefore select None (also see §13.5.1). The middle block of options in the dialog relates to the estimation of the $d$ parameter. It is possible to fix $d$ at zero, in which case the model collapses to an ARMA model. We, however, wish to estimate $d$, and therefore keep the current settings. Finally, we must specify the orders of the process: the values of $p$ and $q$ in ARFIMA($p,d,q$). The highest order both for the autoregressive and for the moving average process is four, so enter that under AR order and MA order. To avoid the estimation of $\phi_1, \phi_2, \phi_3$, click on the dots after Fix AR lags. An edit field will appear, were you should enter 1;2;3 (so no terminating semi-colon). This will enforce $\phi_1 = \phi_2 = \phi_3 = 0$. Do the same for the MA lags, to finish with the completed dialog.\footnote{Unlike X12arima, the current implementation of ARFIMA modelling does not allow for the direct specification of a seasonal ARMA model. A close approximation can be achieved by multiplying out the lag polynomial. For example for the Airline model of §16.2, this would
OK takes us to the model estimation dialog, where an estimation method can be selected, and the sample period adjusted:

Set the estimation method to maximum likelihood. Accepting the dialog, will start the estimation procedure. Despite the complex optimization problem, convergence is rapidly achieved:

```
---- Maximum likelihood estimation of ARFIMA(4,d,4) model ----
The estimation sample is: 1948(3) - 2000 (4)
The dependent variable is: InflatQ
The dataset is: .\OxMetrics\data\UKrpiq.in7

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>d parameter</td>
<td>0.471983</td>
<td>0.03255</td>
<td>14.5</td>
</tr>
<tr>
<td>AR-4</td>
<td>0.897551</td>
<td>0.05747</td>
<td>15.6</td>
</tr>
<tr>
<td>MA-4</td>
<td>-0.763444</td>
<td>0.08015</td>
<td>-9.53</td>
</tr>
<tr>
<td>Constant</td>
<td>4.49412</td>
<td>14.67</td>
<td>0.306</td>
</tr>
<tr>
<td>Seasonal</td>
<td>-0.290712</td>
<td>0.9819</td>
<td>-0.296</td>
</tr>
<tr>
<td>Seasonal_1</td>
<td>2.10488</td>
<td>1.059</td>
<td>1.99</td>
</tr>
<tr>
<td>Seasonal_2</td>
<td>-2.66660</td>
<td>0.9778</td>
<td>-2.73</td>
</tr>
<tr>
<td>D75Q2</td>
<td>13.4495</td>
<td>2.665</td>
<td>5.05</td>
</tr>
<tr>
<td>D79Q3</td>
<td>17.7466</td>
<td>2.665</td>
<td>6.66</td>
</tr>
</tbody>
</table>
```

result in MA terms of order 1, 12, and 13, with $\theta_{13}$ estimated unrestrictedly, instead of imposing $\theta_{13} = -\theta_1 \theta_{12}$. 
Clicking on the Test Summary indicates that the dummies have indeed removed the non-normality (using the test of Doornik and Hansen, 1994):

Descriptive statistics for residuals:
Normality test: Chi^2(2) = 2.0396 [0.3607]
ARCH 1-1 test: F(1, 199) = 3.5010 [0.0628]
Portmanteau(14): Chi^2(11) = 17.216 [0.1016]

There is some evidence for first-order ARCH (indeed, Engle, 1982, used UK inflation to illustrate the ARCH model). The portmanteau test up to lag 14 has 11 degrees of freedom: \( s - (\hat{p} + \hat{q}) - 1 \) when the lag length is \( s \), \( \hat{p} \) and \( \hat{q} \) count the number of ARMA coefficients (but omitting those which were set to zero), and also subtracting one for the estimated \( d \) parameter. This is ignores the presence of non-deterministic regressors, in which case the portmanteau test becomes unreliable.

The estimation results suggest a simplification in the seasonal pattern: the intercept is zero, as is the influence of Q1, but Q2 and Q3 seem to have a cancelling effect (the coefficients almost sum to zero). Similarly, we can combine the two dummies into one variable. Before we impose these four restrictions, we can test them:
which produces:

Test for linear restrictions (Rb=r):

\[
\begin{array}{cccccc}
\text{d parameter} & \text{AR-4} & \text{MA-4} & \text{Constant} & \text{Seasonal} & \text{Seasonal}_1 \\
0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
\text{Seasonal}_2 & \text{D75Q2} & \text{D79Q3} & & & \\
0.0000 & 0.0000 & 0.0000 & & & \\
0.0000 & 0.0000 & 0.0000 & & & \\
1.0000 & 0.0000 & 0.0000 & & & \\
0.0000 & 1.0000 & -1.0000 & & & \\
\end{array}
\]

\[
\begin{array}{c}
r \end{array}
\]

\[
\begin{array}{c}
0.0000 \\
0.0000 \\
0.0000 \\
0.0000 \\
\end{array}
\]

**LinRes Chi^2(4) = 1.48548 [0.8292]**

Therefore we construct:

\[
\text{"Q2-Q3" = (period() == 2) - (period() == 3);}
\]

\[
\text{"D75Q2-D79Q3" = D75Q2 + D79Q3;}
\]

These variables are already in the database.
Re-estimating the modified model leads to:

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>d parameter</td>
<td>0.470532</td>
<td>0.03283</td>
<td>14.3</td>
</tr>
<tr>
<td>AR-4</td>
<td>0.897394</td>
<td>0.05732</td>
<td>15.7</td>
</tr>
<tr>
<td>MA-4</td>
<td>-0.762972</td>
<td>0.08043</td>
<td>-9.49</td>
</tr>
<tr>
<td>Q2-Q3</td>
<td>2.37581</td>
<td>0.4868</td>
<td>4.88</td>
</tr>
<tr>
<td>D75Q2+79Q3</td>
<td>15.5919</td>
<td>1.876</td>
<td>8.31</td>
</tr>
</tbody>
</table>

log-likelihood: -534.651973
no. of observations: 210
no. of parameters: 6

AIC.T: 1081.30395
AIC: 5.14906641

mean(InflatQ): 5.86065
var(InflatQ): 31.3427
sigma: 3.05218
sigma^2: 9.31577

Descriptive statistics for residuals:
Normality test: Chi²(2) = 1.9255 [0.3818]
ARCH 1-1 test: F(1,203) = 3.4523 [0.0646]
Portmanteau(14): Chi²(11)= 14.094 [0.2278]

The entries in the Test menu should again be familiar if you have used other PcGive packages before. From the Test menu select Graphic analysis. As the default Actual and fitted values, Residuals (scaled), and Residual autocorrelations, are marked:

The resulting graphs are in Figure 12.2. The first shows the actual and fitted values, followed by the scaled residuals $\hat{\varepsilon}_t/\hat{\sigma}_e$, the final graph has the residual ACF.
The Test/Test menu allows for more detailed output:

For example, when clicking on Normality test:

Normality test for residuals
Observations 210
Mean 0.076916
Std.Devn. 3.0512
Skewness 0.21978
Excess Kurtosis 0.13584
Minimum -9.4141
Maximum 8.3226

Asymptotic test: \( \chi^2(2) = 1.8521 \) [0.3961]
Normality test: \( \chi^2(2) = 1.9255 \) [0.3818]

Where the reported normality test is that of Doornik and Hansen (1994). Finally, we consider forecasting. From the Test menu, select Forecasting, choosing 8 forecasts:
Figure 12.3 shows the forecasts after selecting error fans instead of the default error bars. We also switched off naive forecasts, although in practice we have found that the difference between the two is very small (but the naive ones are much faster to compute).

It is more common to forecast the year to year quarterly inflation rate. We could have built a model for $i_t^a = 100 \Delta_4 \log(P_t)$, which would have given us those forecasts directly. Casual experimentation suggests that such a model is harder to build. Indeed, if our current model were correct, and since:

$$
\Delta_4 y_t = \Delta y_t + \Delta y_{t-1} + \Delta y_{t-2} + \Delta y_{t-3},
$$

(12.1)

such a model would require the intermediate lags. Constructing the point forecasts for $i_t^a$ can be done from $i_t^q$ as follows:

1. save the forecasts using Test/Store in Database:
2. Next, using copy and paste to add one year of actual data just before the forecast period.

3. From (12.1) we need to construct the four-period moving average. You can do this with the OxMetrics Calculator, or using Algebra:

\[
\text{annualforc} = \text{movingavg}(\text{Forecast}, 3, 0);
\]

Figure 12.4 shows our forecast together with the forecasts from 10 financial institutions. Our forecasts are at the top-end of the range, however, it is likely that the other forecasts were down-weighted to incorporate the expectation of a recession. But, to win a forecast-guru award, it may not be such a bad strategy to be a bit different!

![ARFIMA forecast](image)

**Figure 12.4** Forecasts of annual changes from ARFIMA\((4, d, 4)\) model for UK inflation

To complete this tutorial, we show how forecasts can be made for the levels of the variable. This requires us to work with the unscaled variable \(i = \Delta \log(P_t) = i_t^q/400\). So, create this variable in the database, and re-estimate the model, which, of course, should give the same answers (except for the coefficients on the dummy variable, which are scaled accordingly). Reactivate the Forecast dialog, and expand the Undo data transformations section. Enter the last insample level of LRPI, which can be copied from the database, and check Undo logarithm:
This results in three graphs, see Figure 12.5, where we have added some pre-sample observations to the last two graphs.

Producing level forecasts from log-level forecasts involves an extra forecast bias correction due to the data transformation:

\[ Y_t^f = \exp(y_t^f + \frac{1}{2}e_t^2), \]

where \( e_t^2 \) is the standard error of the forecast error, see e.g. Granger and Newbold (1986, p.311). The error bands are then computed as:

\[ \exp(y_t^f \pm 2e_t), \]

when the critical value for the error bands is set to two.

We leave further exploration to the reader, and finish this chapter with some suggestions. It is of interest to estimate the model in first differences: \( D\text{Inflat}Q \) on \( DQ2-Q3 \)
and \(D_{75Q2-79Q3}\) and forecast the levels (i.e. inflation). The outcomes suggest that we cannot reject \(d = 0.5\) for inflation. This test is works well because for \(D_{InflatQ}\), \(\hat{d}\) is not close to the border of the parameter space, in contrast to the model for \(InflatQ\). To use the weighted maximum likelihood, it is possible to downweight the high inflation period by giving the period 1974(1) – 1984(1) weight 0.5, setting the remainder to unity. This will show a rise in the log-likelihood. To try \(Z\) variables, enter the \(75Q2-79Q3\) dummy variable as such, noting the increase in the log-likelihood.
Chapter 13

ARFIMA Implementation Details

13.1 Introduction

This chapter gives a summary of the implementation details of the procedures in the Arfima package. It complements Doornik and Ooms (2004), while further computational aspects are discussed in Doornik and Ooms (2003).

13.2 The Arfima model

The basic \textbf{ARMA}(p, q) model is

\[ y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q}, \quad t = 1, \ldots, T, \]

assuming either \( \epsilon_t \sim \text{NID}(0, \sigma^2_{\epsilon}) \), or \( \mathbb{E}[\epsilon_t] = 0 \) and \( \mathbb{E}[\epsilon_t^2] = \sigma^2_{\epsilon} \). Using lag polynomials and introducing a mean \( \mu \) we write:

\[ \Phi(L) (y_t - \mu) = \Theta(L) \epsilon_t. \]

With a fractional integration parameter \( d \), the \textbf{ARFIMA}(p, d, q) model is written as

\[ \Phi(L) (1 - L)^d (y_t - \mu) = \Theta(L) \epsilon_t. \quad (13.1) \]

The autocovariance function (ACovF) of a stationary \textbf{ARMA} process with mean \( \mu \):

\[ c(i) = \mathbb{E} [(y_t - \mu)(y_{t-i} - \mu)], \]

defines the variance matrix of the joint distribution of \( y = (y_1, \ldots, y_T)' \):

\[ V[y] = \begin{pmatrix}
    c(0) & c(1) & \cdots & c(T-1) \\
    c(1) & c(0) & & \vdots \\
    c(2) & c(1) & & \vdots \\
    \vdots & \vdots & \ddots & c(1) \\
    c(T-1) & \cdots & c(1) & c(0)
\end{pmatrix} = \mathcal{T} [c(0), \ldots, c(T-1)] = \Sigma, \quad (13.2) \]
which is a Toeplitz matrix, denoted by $T$. Under normality:

$$y \sim N_T(\mu, \Sigma). \quad (13.3)$$

The autocorrelation function, ACF: $c(i)/c(0)$, of a stationary ARMA process is discussed in many textbooks, and readily computed from the $\phi_i$ and $\theta_i$ using the Ox function armapvar. We often work with the autocovariances scaled by the error variance:

$$r = [r(0) \cdots r(T-1)]' = \sigma_e^{-2} [c(0) \cdots c(T-1)]'. \quad (13.3)$$

### 13.2.1 Autocovariance function

An algorithm for the computation of the ACovF of an ARFIMA process is derived in Sowell (1992):

$$c(i) = \sigma_e^2 \sum_{k=-q}^{q} \sum_{j=1}^{p} \psi_k \zeta_j C(d, p + k - i, \rho_j), \quad (13.4)$$

where

$$\psi_k = \sum_{s=|k|}^{q} \theta_s \zeta_{s-|k|}, \quad \zeta_j^{-1} = \rho \left[ \prod_{i=1}^{p} (1 - \rho_i \rho_j) \prod_{m \neq j} (\rho_j - \rho_m) \right], \quad (13.5)$$

and

$$C(d, h, \rho) = \frac{\Gamma(1-2d)}{[\Gamma(1-d)]^2 (1-d)_h} \frac{(d)_h}{\Gamma(1-d)_h} \left[ \rho^{2p} F(d + h; 1 - d + h; \rho) + F(d - h; 1 - d - h; \rho) - 1 \right]. \quad (13.6)$$

Here $\Gamma$ is the gamma function, $\rho_j$ are the roots of the AR polynomial (assumed distinct), and $F(a, 1; c; \rho)$ is the hypergeometric function, see e.g. Abramowitz and Stegun (1970, Ch. 15):

$$F(a, b; c; \rho) = \sum_{i=0}^{\infty} \frac{(a)_i (b)_i \rho^i}{(c)_i i!},$$

where we use Pochhammer’s symbol:

$$(a)_i = a(a + 1)(a + 2) \cdots (a + i - 1), \quad (a)_0 = 1.$$ 

So $(1)_i$ equals $i!$.

In the absence of AR parameters $(13.4)$ reduces to

$$c(i) = \sigma_e^2 \sum_{k=-q}^{q} \psi_k \Gamma(1-2d) \frac{(d)_{k-i}}{[\Gamma(1-d)]^2 (1-d)_{k-i}}.$$ 

1Note the typo in the equation below (8) in Sowell (1992, p.173): $\Gamma(d + s - l)$ in the numerator should read $\Gamma(d - s + l)$. 
13.3 Estimation

13.3.1 Regressors in mean

Any set of exogenous regressors may be used to explain the mean:

\[ z = y - \mu, \quad \mu = f(X, \beta), \]

where \( X \) is a \( T \times k \) matrix. In the leading linear case \( f(X, \beta) = X\beta \) and \( \beta \) is a \( k \times 1 \) vector.

13.3.2 Initial values

Initial values for the parameter estimates are obtained in the order: regressors in mean, \( d \), AR part, and finally MA part. The very first step is to subtract a mean from \( y_t \):

\[ z_t = y_t - \mu_t. \]

When either the sample mean or a specified (known, possibly zero) mean is used: \( \mu_t = \mu \). If regressors are used, take \( \mu_t = f(x_t, \beta) \). In the linear case \( \beta \) is obtained by regression.

1. For the fractional integration parameter the (frequency domain) log periodogram regression of Geweke and Porter-Hudak (1983) is used, yielding \( \hat{d}_0 \). We use \( \lfloor T^{-1/2} \rfloor \) nonzero periodogram points, except when \( p = q = 0 \) when we use all available points. The initial time domain residuals are then obtained using the Ox function \texttt{diffpow}, which computes ‘naive’ fractional differences of \( z_t \):

\[ u_t = \sum_{j=0}^{t} \frac{(-\hat{d}_0)^j}{j!} z_{t-j}. \quad (13.7) \]

2. Next, AR starting values are obtained from solving the Yule-Walker equations taking the number of MA parameters into account:

\[
\begin{pmatrix}
\hat{\rho}(q) & \ldots & \hat{\rho}(q - p + 1) \\
\vdots & & \vdots \\
\hat{\rho}(q + p - 1) & & \hat{\rho}(q)
\end{pmatrix}
\begin{pmatrix}
\phi_0 \\
\vdots \\
\phi_0(p + 1)
\end{pmatrix}
= \begin{pmatrix}
\hat{\rho}(q + 1) \\
\vdots \\
\hat{\rho}(q + p)
\end{pmatrix},
\]

where \( \hat{\rho}(i) \) is the empirical autocorrelation of \( u_t \). When \( q \) is zero, the matrix on the right-hand side is the Toeplitz matrix \( T[\hat{\rho}(0), \ldots, \hat{\rho}(p - 1)]. \)

We use OLS to solve this system, giving the initial estimates of the AR coefficients \( \hat{\phi}_0 \); this will also give a solution when the matrix is singular. Subsequently, the \texttt{arma0} function is used to obtain autoregression residuals \( u_t^\ast \).

3. Starting values for the MA parameters are derived from \( u_t^\ast \) using Tunnicliffe-Wilson’s method, see Granger and Newbold (1986, p.88). Because this iterative method is slow to converge, we choose rather loose convergence criteria. A non-invertible MA is ‘flipped’ to an invertible MA by inverting roots outside the unit circle. The \texttt{arma0} function is then used to obtain ARMA residuals \( u_t^{\ast\ast} \).
When the initial values are used as starting values for further estimation, the following adjustments are made:

1. If \( d \) is not significant at 5%, it is set to zero. A value of \( \hat{d}_0 \) less than \(-0.45\) is set to \(-0.40\), and similarly to \(0.40\) for a value greater than \(0.45\).

2. If \( q > 0 \) and the solution from the Yule-Walker equations yields non-stationary AR parameters, the method is applied as if \( q = 0 \).

### 13.3.3 Exact maximum likelihood (EML)

Based on normality (13.3), and with a procedure to compute the autocovariances in (13.2), the log-likelihood is simply (writing \( z \) for the data vector used for maximization):

\[
\log L (d, \phi, \theta, \beta, \sigma^2) = -\frac{T}{2} \log (2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} z' \Sigma^{-1} z. \tag{13.8}
\]

It is convenient to concentrate \( \sigma^2 \) out of the likelihood, starting by writing \( \Sigma = R\sigma^2 \):

\[
\log L (d, \phi, \theta, \beta, \sigma^2) \propto -\frac{T}{2} \log |R| - \frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} z'R^{-1}z.
\]

Differentiating with respect to \( \sigma^2 \), and solving yields

\[
\hat{\sigma}_e^2 = T^{-1} z'R^{-1} z, \tag{13.9}
\]

with concentrated likelihood (CLF):

\[
\ell_c (d, \phi, \theta, \beta) = -\frac{T}{2} \log (2\pi) - \frac{T}{2} - \frac{1}{2} \log |R| - \frac{T}{2} \log [T^{-1} z'R^{-1} z].
\]

When \( f(X, \beta) = X\beta \) it is more convenient to also concentrate \( \beta \) out of the likelihood. The resulting normal profile log-likelihood function becomes:

\[
\ell_p (d, \phi, \theta) = -\frac{T}{2} (1 + \log 2\pi) - \frac{1}{2} \log |R| - \frac{T}{2} \log [T^{-1} \hat{z}' R^{-1} \hat{z}], \tag{13.10}
\]

where

\[
\hat{z} = y - X\hat{\beta}, \quad \hat{\beta} = (X'R^{-1}X)^{-1} X'R^{-1} y. \tag{13.11}
\]

The function used in the maximization procedure is:

\[
-\frac{1}{2} \left\{ T^{-1} \log |R| + \log \sigma^2 \right\}, \tag{13.12}
\]

from which the value for the log-likelihood (13.10) is easily derived. The computational procedure described in Doornik and Ooms (1999) writes

\[
\sigma^2_e = T^{-1} z'R^{-1} z = T^{-1} e'e,
\]

with \( |R| \) a by-product of the procedure.

Function (13.12) is maximized using BFGS with numerical derivatives. During estimation, stationarity is imposed at each step by rejecting parameter values which have:
13.3 Estimation

- \( d \leq -5 \) or \( d > 0.49999 \);
- \( |\rho_i| \geq 0.9999 \), where \( \rho_i \) are the roots of the AR polynomial.

In addition, the procedure can fail because:
- inability to compute the roots of the AR polynomial;
- \( \rho \zeta \leq 10^{-11} \), this corresponds to multiple roots, see (13.5).

13.3.4 Modified profile likelihood (MPL)

The modified profile log-likelihood, \( \ell_M \), for the regression model with stationary ARFIMA-errors and \( f(X, \beta) = X\beta \):

\[
\ell_M (d, \phi, \theta) = -\frac{T}{2} (1 + \log 2\pi) - \left( \frac{1}{2} - \frac{1}{T} \right) \log \|R\| - \frac{T-k-2}{2} \log \left[ T^{-\frac{1}{2}}Z\'; R^{-\frac{1}{2}}Z \right] - \frac{1}{2} \log \|X'R^{-1}X\|,
\]

(13.13)

see An and Bloomfield (1993), who applied the idea of Cox and Reid (1987) to reduce the bias of the EML estimator due to the presence of unknown nuisance parameters of the regressors.

The residual variance estimator now uses \( T-k \), so that it is unbiased when \( p = q = d = 0 \):

\[
\hat{\sigma}_\epsilon^2 = \frac{1}{T-k} \hat{z}'R^{-\frac{1}{2}}\hat{z}.
\]

(13.14)

13.3.5 Non-linear least squares (NLS)

Defining \( e_t \) as the residuals from applying the ARFIMA(\( p, d, q \)) filter to \( y_t - \mu_t \), the residual variance is:

\[
\sigma_\epsilon^2 = \frac{1}{T-k} \sum_{t=1}^{T} e_t^2.
\]

(13.15)

NLS simply maximizes

\[
f(d, \phi, \theta, \beta) = -\frac{1}{2} \log \left( \frac{1}{T} \sum_{t=1}^{T} e_t^2 \right).
\]

(13.16)

The ARFIMA filter is computed using the Ox function \texttt{diffpow}, see (13.7), followed by \texttt{arma0}. Since (13.7) essentially drops the first observation, \( e_1 = 0 \) when \( d \) is estimated.

Function (13.16) is maximized using BFGS with numerical derivatives, optionally with stationarity imposed.

13.3.6 Variance-covariance matrix estimates

Let \( \vartheta' = [d' \phi' \theta'] \). The variance-covariance matrix for the EML (\( \ell = \ell_P \)) and MPL (\( \ell = \ell_M \)) estimates is computed as:

\[
\begin{pmatrix}
-\partial^2 \ell (\vartheta) / \partial \vartheta \partial \vartheta' |_{\hat{\vartheta}} & 0 \\
0 & \hat{\sigma}_\epsilon^2 \left( X'R^{-1}X \right)^{-1}
\end{pmatrix}
\]

\[ \left. \begin{pmatrix}
-\partial^2 \ell (\vartheta) / \partial \vartheta \partial \vartheta' |_{\hat{\vartheta}} & 0 \\
0 & \hat{\sigma}_\epsilon^2 \left( X'R^{-1}X \right)^{-1}
\end{pmatrix} \right. \]
The second derivative of $\ell$ is computed numerically.

For NLS, the variance-covariance is the inverse of minus the numerical second derivative of (13.16).

### 13.4 Estimation output

Estimation output consists of

- Estimated coefficients, with estimated standard errors, $t$-values, and $p$-values. The $p$-values are based on a $t(T - s)$-distribution, where $s$ is the number of estimated parameters, including the residual variance. When all parameters are freely estimated: $s = 1 + p + q + k + 1$.
- log-likelihood $\hat{\ell}$:
  - EML: $\ell_c$,
  - MPL: $\ell_M$,
  - NLS: $f - \frac{T}{2} (1 + \log 2\pi)$,

where $f$ for NLS is from (13.16). So at the same parameter values, EML and NLS would report the same log-likelihood.
- Akaike information criterion

$$\text{AIC.T} = -2\hat{\ell} + 2s,$$

where $s$ the number of estimated parameters. When no parameters are fixed: $s = 1 + p + q + k + 1$ (the last accounts for the residual variance). The AIC.T/T is also reported.
- Residual variance: (13.9) for EML, (13.14) for MPL, (13.15) for NLS.
- Mean and variance of dependent variable.
- BFGS convergence criteria, convergence result and starting values.

### 13.5 Estimation options

#### 13.5.1 Sample mean versus known mean

It has been found in early Monte Carlo experiments that, in smaller samples, using the theoretical mean could lead to more accurate estimation of $d$ (see e.g. Cheung and Diebold, 1994). This can be seen as the most effective way to reduce the effect of a nuisance parameter on inference for the parameter of interest. Therefore, the Arfima package allows for fixing the mean at a specified value. Let $y_t$ denote the original dependent variable, and $\mu_y$ the known mean.

The $z$ used in §13.3.3 for estimation when specifying a known mean is:

$$z_t = y_t - \mu_y,$$

otherwise the package uses

$$z_t = y_t - \hat{\mu}_y.$$
The specification of the mean affects the likelihood. For the last term in the log-likelihood:

\[(y - \mu)' R^{-1} (y - \mu) = y' R^{-1} y - 2\mu' R^{-1} y + \mu' R^{-1} \mu,\]

so the known mean case adds \(\mu' R^{-1} y\), whereas the second case adds \(\hat{\mu}' R^{-1} y\), and different results must be expected.

### 13.5.2 Fixing parameters

It is possible to fix \(d\) at a specific value, or drop ARMA terms.

### 13.5.3 Weighted estimation

A weight variable \(w\), \(w_t \geq 0\), can be used in estimation. Write \(\bar{w}_t = w_t/\bar{w}_{>0}\), where \(\bar{w}_{>0}\) is the mean of the positive weights.

Then (13.12) for EML becomes:

\[-\frac{1}{2} \left\{ T^{-1} \log |R| - T^{-1} \sum_{\bar{w}_{>0}} \log \bar{w}_t + \log T^{-1} \sum_{t=1}^T e_t^2 \bar{w}_t \right\},\]

The NLS function is adjusted in a similar fashion. Weighted estimation is not available for MPL, and weights are ignored for forecasting. The weighted residuals, \(\hat{e}_t \bar{w}_t^{-1/2}\), are used in the residual-based diagnostics.

### 13.5.4 \(Z\) variables

With both additive normal regressors \(x_t\) and innovation \(Z\) variables \(z_t\) the ARFIMA model becomes:

\[\Phi (L) (1 - L)^d (y_t - x_t' \beta) = \Theta (L) (\varepsilon_t + z_t' \gamma).\]

The notation for the \(Z\) variables in this subsection should not be confused with \(z_t\), the demeaned \(y_t\). After applying the normal EML or NLS filter to \(z_t\), \(z_t' \hat{\gamma}\) is subtracted at each iteration.

This model has the familiar ADL (Autoregressive Distributed-lag model) as a special case, since \(z_t\) can contain different lags of the same (exogenous) variable. Whereas additive outliers (and missing observations) can be estimated using dummies for the \(X\) variables, see e.g. Brockwell and Davis (1993, §12.3), innovation outliers can be modelled by dummies for \(Z\) variables. Note that adding a single observation dummy for a \(Z\) variable has the same effect as giving that observation zero weight in the \(W\) variable. This is illustrated in fracest8.ox.

\(Z\) variables are not available for MPL.
13.6 Forecasting

Two methods of forecasting are supported, based on the results in Beran (1994, §8.7). As before let $z = (z_1, \ldots, z_T)'$ denote the observations over the estimation period. Assume $z_t$ is stationary and $d > -1$. The best linear prediction of $z_{T+h}$ is

$$\hat{z}_{T+h} = \left[ r(T-1+h) \cdots r(h) \right] \{ T \left[ r(0), \ldots, r(T-1) \right] \}^{-1} z = q'z,$$

which consists of the reversed ACovF starting from $h$, times the original data weighted by their correlations. The `solveCovF` function is used to solve $Tx = z$ in order to keep storage requirements of order $T$, see Doornik and Ooms (1999). The mean square error is

$$\text{MSE}(\hat{z}_{T+h}) = \sigma^2 (r(0) - r'q).$$

In the presence of a mean-function $\mu_t=f(x_t, \beta)$ the forecasts are:

$$\hat{y}_{T+h} = q' (y - \mu) + \mu_{t+h} + x'_{T+h} \hat{\beta}.$$

The program computes all requested forecasts $\hat{z}_h = (z_{T+1}, \ldots, z_{T+h})'$ and their joint variance-covariance matrix, $\text{Cov} (\hat{z}_h)$ simultaneously. $\text{Cov} (\hat{z}_h)$ is also used to derive the mean squared errors for partial sums, $\sum_{i=1}^{h} \hat{z}_{T+i}$, integrated partial sums etc.

‘Naive’ forecasts are derived from the autoregressive representation of the process, truncated at $T+h$:

$$\Theta^{-1} (L) \Phi (L) (1 - L)^d z_t = (1 - b_1 L \cdots - b_{T+h-1} L^{T+h-1}) z_t = B(L) z_t.$$ 

In this case the $z_t$ need not be stationary, cf. Beran (1995), but $d > -0.5$. The first $T$ coefficients in the $(1 - L)^d$ polynomial can be computed using the `diffpow` function when the input is a one followed by $T-1$ zeroes; this follows from (13.7). For polynomial multiplication and division, `polymul` and `polydiv` are used. The naive forecasts are computed recursively:

$$\hat{z}_{T+h}^* = [b_{T+h-1} \cdots b_1] \times [z' \; \hat{z}_{T+1} \cdots \hat{z}_{T+h-1}]',$$

and

$$\text{MSE}(\hat{z}_{T+h}^*) = \sigma^2 \left( 1 + \sum_{i=1}^{h-1} a_i^2 \right), \quad (13.17)$$

where $a_i$ are the coefficients of $B^{-1} (L)$.

Level forecasts are computed by adding the (integrated) partial sums of forecasts to a specified starting level. The reported MSE of the integrated naive forecasts can be obtained directly from (13.17).

Forecasting with Z variables is not yet implemented.
13.7 ARFIMA batch commands

ARFIMA \((p, d, q)\);
ARIMA \((p, d, q)\);
ARM \((p, q)\);

Specifies the model to estimate, for example ARFIMA \((1, d, 1)\) to estimate \(d\) with AR(1) and MA(1) components; ARFIMA \((1, 0.3, 1)\) to fix \(d = 0.3\), or ARMA \((1, 1)\) to fix \(d = 0\).

fixAR(fix1, fix2, ...);
fixMA(fix1, fix2, ...);

Fixes the specified AR or MA parameters.

fixmean(fixat);

Fixes the mean at the specified value.

samplemean;

Estimates in deviation from sample mean.

forecast(ctforc);

prints ctforc forecasts.

estimate("method"=OLS, year1=-1, period1=0, year2=-1, period2=0);

Estimate a system. The presence of default arguments implies that the shortest version is just: estimate(), which estimates by OLS using the maximum possible sample, and no forecasts. Similarly, a call to estimate("EML","1950",1) corresponds to estimate("EML","1950",1,-1,0).

The method argument is one of:

<table>
<thead>
<tr>
<th>method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EML</td>
<td>exact maximum likelihood,</td>
</tr>
<tr>
<td>NLS</td>
<td>non-linear least squares,</td>
</tr>
<tr>
<td>MPL</td>
<td>modified profile likelihood,</td>
</tr>
<tr>
<td>Init</td>
<td>Initialization (starting values) only,</td>
</tr>
<tr>
<td>NLS-STAT</td>
<td>non-linear least squares with stationarity imposed</td>
</tr>
</tbody>
</table>

year1(period1) – year2(period2) is the estimation sample. Setting year1 to \(-1\) will result in the earliest possible year1(period1), setting year2 to \(-1\) will result in the latest possible year2(period2).

If the database is dated, specify the start date in the year1 argument and end date in year2, with both periods set to zero. An example is estimate("EML","1984-01-04",0,1992-01-02,0).

option("option", argument);
module("PcGive");

Starts the PcGive module. If PcGive is already running, this batch command is not required.

package("ARFIMA");

Use this command to select the ARFIMA package.

progress;

Reports the modelling progress.

store("name", "rename"="");

Use this command to store residuals, etc. into the database, the default name is used. Note that if the variable already exists, it is overwritten without warning. The name must be one of:
- residuals
- forecasts
- forcse
- forclevel
- forclevelse

The optional second argument replaces the default name. For example store("residuals") stores the residuals under the name Residual; store("residuals", "xyz") stores them under the name xyz.

system { Y=...; X=...; Z=...; W=...; }

Specify the system, consisting of the following components:
- Y endogenous variable;
- X non-modelled X variable;
- Z non-modelled Z variable;
- W weight variable.

The variables listed are separated by commas, their base names (that is, name excluding lag length) must be in the database. If the variable names are not a valid token, the name must be enclosed in double quotes.

The following special variables are recognized: Constant, Trend, Seasonal and CSeasonal.

test("test", lag1=0, lag2=0);

Performs a specific test using the specified lag lengths.
"ar" test for autocorrelated errors from lag1 to lag2;
"arch" ARCH test up to order lag1;
"normal" normality test;
"rescor" residual correlogram up to lag lag1;

testlinres {...}
Test for linear restrictions. The content is the matrix dimensions followed by the (R : r) matrix.
testsummary;
Do the test summary.
Part VI

X12arima for OxMetrics
Chapter 14

Overview of X12arima for OxMetrics

14.1 Introduction

The X12arima for OxMetrics program gives access to most of the X-12-ARIMA features for seasonal adjustment and regression with ARIMA errors (regARIMA) modelling. Variables are selected from a OxMetrics database, and an X12 specification is formulated using dialogs. After successful estimation, the results can be graphed, or stored in a OxMetrics database. OxMetrics graphs can be manipulated on screen, and saved to a variety of formats.

14.2 X-12-ARIMA

X-12-ARIMA is the outcome of a research programme on seasonal adjustment which has been undertaken by researchers at the US Census Bureau since the 1950’s. At the heart is the X-11 seasonal adjustment method, which is widely used by statistical agencies throughout the world.1 X-12-ARIMA builds on X-11-ARIMA, adding the ability to extend the time series with forecasts and backcasts from regARIMA models prior to seasonal adjustment.

The current version, called X12arima for OxMetrics, uses the code created and made available by the Census Bureau, with minor modifications. The Census version works through a specification file (.spc), which sets the options and determines which actions to take. The OxMetrics version preserves the specification syntax, but builds a specification through a set of dialogs. It is not necessary to first create .spc files: seasonal adjustment models can be formulated entirely interactively, with the data loaded in OxMetrics.2

1 Other techniques for seasonal adjustment are incorporated in the STAMP program, see Koopman, Harvey, Doornik, and Shephard (1999), and Tramo/Seats, see www.bde.es.
2 The dialogs only allow a subset of the available options that can be set in .spc files. An Ox package will be available that provides access to the full range of options.
14.3 Credits


This version of X-12-ARIMA includes an automatic ARIMA model selection procedure based largely on the procedure of Gomez and Maravall as implemented in TRAMO.

Primary Programmers: Brian Monsell, Mark Otto

We suggest that you cite Findley, Monsell, Bell, Otto, and Chen (1998) whenever reporting results from X12arima for OxMetrics.

14.4 Disclaimer

The X-12-ARIMA program is made available for evaluation purposes only. Therefore this also holds for X12arima for OxMetrics. No warranty whatsoever is given for these programs.

14.5 Limitations

X12arima for OxMetrics is an easy-to-use program, but there are some limitations regarding its use:

- X12arima for OxMetrics requires a valid PcGive registration to work.
- X12arima for OxMetrics is primarily for quarterly and monthly data.
- The sample size and number of regressors is restricted, as is the number of forecasts:

<table>
<thead>
<tr>
<th>32-bit</th>
<th>64-bit</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td>300</td>
</tr>
<tr>
<td>600</td>
<td>3000</td>
</tr>
<tr>
<td>70</td>
<td>250</td>
</tr>
<tr>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

- There is no online Help.

14.6 Documentation

Consult www.census.gov/srd/www/x12a/ to download the full X-12-ARIMA documentation, in particular:

- X-12-ARIMA Reference Manual
- X-12-ARIMA Quick Reference
Another useful source of information is Ladiray and Quenneville (2001), who provide a detailed study of the X-11 method, thereby updating Shiskin, Young, and Musgrave (1967).

### 14.7 Census X-11 Seasonal Adjustment

The Census X-11 is the main building block in the X-12-ARIMA procedure. The objective is to decompose a monthly or quarterly time series \( X_t \) into a trend component \( T_t \), a seasonal component \( S_t \), and a residual or irregular component \( I_t \). Especially in monthly data, there can also be important effects caused by the calendar (e.g. the timing of moving holidays, the number of trading days in a month), denoted by \( D_t \). The available types of decomposition are:

- **multiplicative**
  \[ X_t = T_t S_t I_t D_t, \]

- **additive**
  \[ X_t = T_t + S_t + I_t + D_t, \]

- **log-additive**
  \[ \log Y_t = \log T_t + \log S_t + \log I_t + \log D_t, \]

- **pseudo-additive**
  \[ X_t = T_t (S_t + I_t + D_t). \]

According to Findley, Monsell, Bell, Otto, and Chen (1998), the most frequently used decomposition is the multiplicative version.

#### Table 14.1 Summary of steps taken in the Census X-11 method

<table>
<thead>
<tr>
<th>part</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Optional prior adjustments for user-specified calendar effects</td>
</tr>
<tr>
<td>B</td>
<td>Preliminary estimation of trading-day variation and irregular</td>
</tr>
<tr>
<td>C</td>
<td>Final estimation of trading-day variation and irregular</td>
</tr>
<tr>
<td>D</td>
<td>Final estimation of seasonal factors, trend-cycle and irregular</td>
</tr>
<tr>
<td>E</td>
<td>Modification for extreme observations</td>
</tr>
<tr>
<td>F</td>
<td>Report of summary measures</td>
</tr>
</tbody>
</table>

Table 14.1 summarizes the steps taken in the Census X-11 method. We now consider these in more detail, listing the primary output generated in X12arima for OxMetrics (all the output appears in tables, which are identified by the part and a sequence number):

**A** Optionally, and before any seasonal adjustment is made, various prior adjustments can be made from user-supplied factors. RegARIMA models can be used to create forecasts that are used to extend the series prior to seasonal adjustment.

**A1** unadjusted original series A1.

**B** Preliminary trading-day adjustment factors and weights for reducing the effect of extreme and near-extreme irregular values.

**B1** original series, or prior adjusted series from part A

**B17** Preliminary weights for irregular component
Chapter 14 Overview of X12arima for OxMetrics

C The original series adjusted for trading-day variation is modified for extreme and near-extreme values with the B17 weights, and improved trend-cycle and seasonal estimates are obtained.

C17 Final weights for irregular component

D The original series adjusted for trading-day variation is modified for extreme and near-extreme values with the C17 weights. If adjustment for regARIMA seasonal was used, this is combined with the X-11 seasonal. This results in the final trend-cycle and seasonal estimates.

D10 Final seasonal factors
D11 Final seasonally adjusted series
D12 Final trend cycle
D13 Final irregular

E The original and seasonally adjusted series are modified for extremes. Tables are provided to assess the quality of the seasonal adjustment.

E1 Modified original series, modified for extremes
E2 Modified seasonally adjusted series, modified for extremes
E3 Modified irregular series, modified for extremes

F Months for cyclical dominance (MCD) moving average and summary measures
F3 Monitoring and quality assessment statistics

14.8 X-12-ARIMA Seasonal Adjustment

The X-12-ARIMA procedure enhances the Census X-11 method by extending the series with forecasts and backcasts from a model for the series. This has been found generally to lead to smaller revisions as new observations are added.

This is incorporated through optional pre-estimation of a regARIMA model (linear regression model with ARIMA time series errors, see §14.9). This model provides forecasts (and optionally backcasts) and prior adjustments for various effects before steps B and C of the X-11 method are applied.

14.9 regARIMA

The basic ARMA$(p, q)$ model in regARIMA is

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}, \quad t = 1, \ldots, T,$$

assuming either $\varepsilon_t \sim \text{NID}[0, \sigma_\varepsilon^2]$, or $E[\varepsilon_t] = 0$ and $E[\varepsilon_t^2] = \sigma_\varepsilon^2$. Using $L$ for the lag operator, $L y_t = y_{t-1}$, lag polynomials $\Phi, \Theta$, and introducing a mean $\mu$:

$$\Phi(L) (y_t - \mu) = \Theta(L) \varepsilon_t.$$

The ARIMA model corresponds to an ARMA model applied to the $d$th differences of $y_t$, i.e. to $(1 - L)^d = \Delta^d y_t$. Most commonly $d = 1$, The regARIMA model allows for
seasonal specification and regressors in the mean:

\[ \Phi_p(L)\Phi_P(L^s) (1 - L)^d (1 - L^s)^D (y_t - \mathbf{x}_t'\mathbf{\beta}) = \Theta_q(L)\Theta_Q(L^s)\varepsilon_t, \]  

(14.1)

where \( s \) is the seasonal frequency, and subscripts on the lag polynomial indicate the order. Short-hand notation for this is to write ARIMA\((p, d, q)(P, D, Q)\), or, when the seasonal part is absent: ARIMA\((p, d, q)\). For example, an ARIMA\((1, 0, 1)(1, 0, 1)\) is:

\[ (1 - \phi L) (1 - \phi_4 L^4) y_t = (1 - \theta L) (1 - \theta_4 L^4) \varepsilon_t, \]

or equivalently:

\[ y_t^* = \phi y_{t-1} + \phi_4 y_{t-5} - \phi \varepsilon_{t-1} - \theta_4 \varepsilon_{t-4} + \theta_4 \varepsilon_{t-5} + \varepsilon_t. \]

The frequency subscript on the seasonal part is omitted when it is implied from the frequency of the data. So the airline model can also be written as an \((0 1 1)(0 1 1)\) model, which for monthly data amounts to:

\[ (1 - L) (1 - L^{12}) y_t = (1 - \theta L) (1 - \theta_1 L^{12}) \varepsilon_t. \]

Regressors enter in the mean, i.e. before differencing and before the autoregressive polynomials are applied. As a consequence, when \( d = 1 \) the constant drops out (is differenced to zero), when \( d = 2 \) the constant and trend drop out, etc. X12-ARIMA has the option to add a so-called trend-constant \( c_t \) which has order offsetting \( d + D \):

\[ (1 - L)^d (1 - L^s)^D c_t = I(t \geq 1), \]

where \( I \) is the indicator function. If such a variable is included, (14.1) can be written as:

\[ \Phi_p(L)\Phi_P(L^s) (1 - L)^d (1 - L^s)^D \left( y_t - \mathbf{x}_t'\mathbf{\beta} \right) - \mu = \Theta_q(L)\Theta_Q(L^s)\varepsilon_t. \]

Note that

1. \textit{regARIMA} reports the MA coefficients with the opposite sign from ARFIMA, cf. (11.1);
2. unlike in dynamic regression models, the regressors enter inside the lag polynomial.

The \textit{regARIMA} procedure allows for modelling along the lines of the Box–Jenkins method (see, Box, Jenkins, and Reinsel, 1994), including automatic model selection and outlier detection. It is possible to estimate ARIMA models without performing any seasonal adjustment.

### 14.10 ARIMA model selection

X12-ARIMA has three methods to specify the ARIMA component:

1. directly specify the ARIMA form \((p d q)(P D Q)\);
3. Automatic selection based as used in previous versions of X12-ARIMA (\texttt{pickmdl} command, which was called \texttt{automdl} in version 0.2.9 as used in X12arima for GiveWin).
14.11 X12arima in PcGive

The X12arima procedure that is part of PcGive provides a convenient way to apply a simplified version of X12-ARIMA to one or more series. However, this method of operation is not useful for routine seasonal adjustment of many series in a production environment.

Transformation
- Logarithms or levels;
- Differencing, \( d = 0, 1, 2 \): none, \( \Delta, \Delta^2 \);
- Seasonal differencing, \( D = 0, 1 \): none or \( \Delta_s = 1 - L_s \).

ARMA model specification
- No ARIMA errors; estimates regression if any regressors are specified, otherwise there is no model estimation at all;
- estimate \((p d q)(P D Q)\) as specified;
- Automatic ARMA selection using the TRAMO approach, with \( d, D \) fixed as specified above;
- Automatic ARIMA selection using the TRAMO approach, including automatic selection of \( d, D \).

Regression
- User-defined regressors Database variables that were added to the selection as X variables will be included as user-defined regressors (and may impact on the number of possible forecasts).
- Intercept after differencing A trend-constant, see §14.9, can be added by selecting ‘intercept after differencing.’ This amounts to a trend for first differences, and a quadratic trend for second differences.
- Fixed seasonals Fixed seasonals can be added, either as \( s - 1 \) (differenced) seasonal dummies or \( s - 1 \) sine-cosine terms. When fixed seasonals are present, the seasonal ARIMA part \((P, D, Q)\) is automatically omitted.
- Trading day effects for flows Trading day adjustments capture calendar effects for flow variables:
  - Trading days (six regressors):
    \[
    \text{(no. of } w \text{)} - \text{(no. of Sundays)}, w = \text{Mon, Tue, ..., Sat}
    \]
  - Weekend adjustment (one regressors):
    \[
    \text{(no. of weekdays)} - \frac{5}{2}\text{(no. of Saturdays and Sundays)}
    \]
  - Leap year adjustment:
No transformation: Leap year variable (one regressors):

\[0.75I(\text{leap-year February}) - 0.25I(\text{non-leap February}),\] monthly,

\[0.75I(\text{leap-year first quarter}) - 0.25I(\text{non-leap first quarter}),\] quarterly.

Transformation is logarithms: Leap year scale factor is applied to the dependent variable:

\[
\frac{28.25}{\text{no. of days in February}} I(\text{February}) + I(\text{non-February}), \quad \text{monthly},
\]

\[
\frac{31 + 28.25 + 31}{\text{no. of days in Q1}} I(\text{Q1}) + I(\text{non Q1}), \quad \text{quarterly}.
\]

In that case, all months of February (or first quarters) are rescaled.

- Length of month/quarter (one regressors):

  (no. of days in month) − 30.4375, \quad \text{monthly},

  (no. of days in quarter) − 91.3125, \quad \text{quarterly}.

Outlier detection

Three types of outliers can be detected automatically:

- **Additive outlier (ao)** is a dummy variable for one observation:

  \[AO_{t_0}^{(t)} = I(t = t_0).\]

- **Level shift (ls)** is a step dummy which is −1 prior to date \(t_0\) when it changes permanently to zero:

  \[LS_{t_0}^{(t)} = I(t \geq t_0) − 1.\]

- **Temporary change (tc)** is a step dummy which is zero prior to date \(t_0\) when it changes to 1 with subsequent exponential decay:

  \[TC_{t_0}^{(t)} = \alpha^{t - t_0} I(t \geq t_0).\]

So from the break onwards, the variable takes the values 1, \(\alpha\), \(\alpha^2\), .... The default rate of decay is set to:

\[\alpha = 0.7^{12/s},\]

which is 0.7 for monthly data and 0.343 for quarterly data.

Seasonal adjustment

Specify whether to run X11 seasonal adjustment after regARIMA.

If the transformation is in logarithms, the multiplicative mode is used. Additive mode is used for levels.
Chapter 15

Tutorial on Seasonal Adjustment with 
X12arima for OxMetrics

15.1 Introduction

The X12arima for OxMetrics program gives access to some of the X-12-ARIMA features for seasonal adjustment and regARIMA modelling. Variables are selected from a OxMetrics database, and an X12 specification is formulated using dialogs. After successful estimation, the results can be graphed, or stored in a OxMetrics database. OxMetrics graphs can be manipulated on screen, and saved to a variety of formats.

15.1.1 Loading the first data set

The first step is to load data into OxMetrics which X12arima can then access. Here we start with a data set called ICMETI.in7. This can be found in the X12arima folder of OxMetrics7.

The .in7 extension indicates an OxMetrics data file. The .in7 file is a text file, describing the data. There is a companion .bn7 file, which holds the actual numbers (in binary format). OxMetrics can handle a wide range of data files: details are in the OxMetrics book.

The ICMETI variable in ICMETI.in7 holds the monthly series entitled ‘Total Inventories Communications Equipment’ from 1968M1 to 1989M3. After loading the data, create a graph of the ICMETI series, as in Figure 15.1.

15.1.2 Quick Seasonal Adjustment

Start with Formulate to select ICMETI. Make sure the ICMETI database is selected, and double click on the ICMETI variable to add it to the list of selected variables (or click once to select it, and then use the << button):
Press OK to go to the Model Settings dialog, keeping the default settings:

- Transformation: logarithms
- ARMA: automatic ARIMA
  This also selects the level of differencing, so the differencing choice under transformation is ignored.
- Regression: nothing selected
Seasonal adjustment: check the X11 option
Because the logarithmic transformation is used, the seasonal adjustment mode of X11 is multiplicative.

Press OK to go to the Estimate dialog. Press OK again to accept the default full sample estimation. A graphics window with name X12-ICMETI is created automatically, as in Figure 15.2.

(a) Time-series plot of the dependent variable and its seasonally adjusted version, labelled SA.

(b) Forecasts of the dependent variable, derived from the estimated regARIMA model. The forecasts are shown with one standard-error bands (68% coverage).

(c) The irregular from stage D of the X11 procedure.

(d) Spectrum of the detrended (differenced) dependent variable and of the seasonally adjusted variable. The seasonal frequencies are marked with a circle, while the trading-day frequencies are marked with a cross.

The output from each run appears in the Results window inside OxMetrics, in compact format (more extensive output can be requested under Options):

U. S. Department of Commerce, U. S. Census Bureau

X-12-ARIMA monthly seasonal adjustment Method,
Release Version 0.3  Build 192
This method modifies the X-11 variant of Census Method II
15.1 Introduction

and the X-11-ARIMA program based on the methodological research developed by Estela Bee Dagum, Chief of the Seasonal Adjustment and Time Series Staff of Statistics Canada, September, 1979.

This version of X-12-ARIMA includes an automatic ARIMA model selection procedure based largely on the procedure of Gomez and Maravall (1998) as implemented in TRAMO (1996).

Primary Programmers: Brian Monsell, Mark Otto

Series Title- X-12-ARIMA run of ICMETI
Series Name- ICMETI

-Period covered- 1st month, 1968 to 3rd month, 1989
-Type of run - multiplicative seasonal adjustment
-Sigma limits for graduating extreme values are 1.5 and 2.5.
-3x3 moving average used in section 1 of each iteration,
-3x5 moving average in section 2 of iterations B and C,
-moving average for final seasonal factors chosen by Global MSR.

Specifications:

Line #
-----
1: # spec file generated by Ox
2: series
3: {
4:   file="ox"
5: }

Figure 15.2  Seasonally adjusted ICMETI
6: transform
7: {
8:   function=log
9: }
10: forecast
11: {
12:   probability=0.682689
13: }
14: automdl
15: {
16: }
17: x11
18: {
19: }

Automatic ARIMA Model Selection

"Automatic Modeling Methods for Univariate Series",
A Course in Time Series
(Edited by D. Pena, G. C. Tiao, R. S. Tsay),
New York : J. Wiley and Sons

Maximum order for regular ARMA parameters :  2
Maximum order for seasonal ARMA parameters :  1
Maximum order for regular differencing :  2
Maximum order for seasonal differencing :  1

Results of Unit Root Test for identifying orders of differencing:
Regular difference order :  2
Seasonal difference order :  0

Mean is not significant.

Automatic model choice : (2 2 2)(1 0 1)

Final automatic model choice : (2 1 1)(1 0 1)
End of automatic model selection procedure.
Estimation converged in  16 ARMA iterations, 116 function evaluations.

Regression Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>.0068</td>
<td>.00518</td>
<td>1.32</td>
</tr>
</tbody>
</table>

ARIMA Model: (2 1 1)(1 0 1)
Nonseasonal differences: 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonseasonal AR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 1</td>
<td>.5511</td>
<td>.13111</td>
</tr>
</tbody>
</table>
Lag 2  .2352  .07705
Seasonal AR
   Lag 12  .8998  .03926
Nonseasonal MA
   Lag 1  .4622  .13061
Seasonal MA
   Lag 12  .6805  .06322
Variance  .17739E-03
SE of Var  .15741E-04

Likelihood Statistics
Number of observations (nobs)  255
Effective number of observations (nefobs)  254
Number of parameters estimated (np)  7
Log likelihood  733.9873
Transformation Adjustment  -2133.5851
Adjusted Log likelihood (L)  -1399.5978
AIC  2813.1956
AICC (F-corrected-AIC)  2813.6509
Hannan Quinn  2823.1568
BIC  2837.9570

Roots of ARIMA Model

<table>
<thead>
<tr>
<th>Root</th>
<th>Real</th>
<th>Imaginary</th>
<th>Modulus</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonseasonal AR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root 1</td>
<td>1.2000</td>
<td>.0000</td>
<td>1.2000</td>
<td>.0000</td>
</tr>
<tr>
<td>Root 2</td>
<td>-3.5433</td>
<td>.0000</td>
<td>3.5433</td>
<td>.5000</td>
</tr>
<tr>
<td>Seasonal AR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root 1</td>
<td>1.1114</td>
<td>.0000</td>
<td>1.1114</td>
<td>.0000</td>
</tr>
<tr>
<td>Nonseasonal MA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root 1</td>
<td>2.1635</td>
<td>.0000</td>
<td>2.1635</td>
<td>.0000</td>
</tr>
<tr>
<td>Seasonal MA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root 1</td>
<td>1.4695</td>
<td>.0000</td>
<td>1.4695</td>
<td>.0000</td>
</tr>
</tbody>
</table>

FORECASTING
Origin 1989.Mar
Number 12

Confidence intervals with coverage probability ( .68269)
On the Original Scale

<table>
<thead>
<tr>
<th>Date</th>
<th>Lower</th>
<th>Forecast</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989.Apr</td>
<td>9165.66</td>
<td>9288.70</td>
<td>9413.39</td>
</tr>
<tr>
<td>1989.May</td>
<td>9131.75</td>
<td>9313.77</td>
<td>9499.42</td>
</tr>
<tr>
<td>1989.Jun</td>
<td>9057.96</td>
<td>9305.53</td>
<td>9559.86</td>
</tr>
</tbody>
</table>
Chapter 15 Tutorial on Seasonal Adjustment with X12arima for OxMetrics

1989.Jul  9048.06  9361.40  9685.59
1989.Aug  9009.25  9387.25  9781.10
1989.Sep  8920.35  9359.21  9819.66
1989.Oct  8895.91  9397.01  9926.34
1989.Nov  8874.04  9435.97  10033.48
1989.Dec  8725.21  9337.39  9992.53
1990.Jan  8732.23  9403.24  10125.80
1990.Feb  8766.16  9496.90  10288.56
1990.Mar  8751.07  9536.16  10391.70

D 8.A F-tests for seasonality (ICMETI)
Test for the presence of seasonality assuming stability.

<table>
<thead>
<tr>
<th></th>
<th>Sum of squares</th>
<th>Dgrs.freedom</th>
<th>Mean square</th>
<th>F-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between months</td>
<td>111.7761</td>
<td>11</td>
<td>10.16146</td>
<td>14.092**</td>
</tr>
<tr>
<td>Residual</td>
<td>175.2209</td>
<td>243</td>
<td>.72107</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>286.9970</td>
<td>254</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Seasonality present at the 0.1 per cent level.

Nonparametric Test for the Presence of Seasonality Assuming Stability

<table>
<thead>
<tr>
<th>Kruskal-Wallis statistic</th>
<th>Dgrs.freedom</th>
<th>Probability level</th>
</tr>
</thead>
<tbody>
<tr>
<td>93.7568</td>
<td>11</td>
<td>.000%</td>
</tr>
</tbody>
</table>

Seasonality present at the one percent level.

Moving Seasonality Test

<table>
<thead>
<tr>
<th></th>
<th>Sum of squares</th>
<th>Dgrs.freedom</th>
<th>Mean square</th>
<th>F-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Years</td>
<td>11.9210</td>
<td>20</td>
<td>.596052</td>
<td>1.779</td>
</tr>
<tr>
<td>Error</td>
<td>73.6957</td>
<td>220</td>
<td>.334980</td>
<td></td>
</tr>
</tbody>
</table>

Moving seasonality present at the five percent level.

COMBINED TEST FOR THE PRESENCE OF IDENTIFIABLE SEASONALITY

IDENTIFIABLE SEASONALITY PRESENT

F 3. Monitoring and Quality Assessment Statistics
The measures below are between 0 and 3; acceptance region from 0 to 1.

1. The relative contribution of the irregular over three months span (from Table F 2.B).
   \[ M_1 = .206 \]

2. The relative contribution of the irregular component to the stationary portion of the variance (from Table F 2.F).
   \[ M_2 = .010 \]

3. The amount of month to month change in the irregular component as compared to the amount of month to month change in the trend-cycle (from Table F2.H).
   \[ M_3 = .000 \]

4. The amount of autocorrelation in the irregular as described by the average duration of run (Table F 2.D).
   \[ M_4 = 1.060 \]

5. The number of months it takes the change in the trend-cycle to surpass the amount of change in the irregular
   \[ M_5 = .000 \]
15.1 Introduction

6. The amount of year to year change in the irregular as compared to the amount of year to year change in the seasonal (from Table F 2.H).

7. The amount of moving seasonality present relative to the amount of stable seasonality (from Table F 2.I).

8. The size of the fluctuations in the seasonal component throughout the whole series.

9. The average linear movement in the seasonal component throughout the whole series.

10. Same as 8, calculated for recent years only.

11. Same as 9, calculated for recent years only.

*** ACCEPTED *** at the level .50

*** Check the 4 above measures which failed.

*** Q (without M2) = .56 ACCEPTED.

15.1.3 Test menu

More graphicl options are available under Graphic Analysis on the Test menu:

![Image of Test menu dialog]

For example, this dialog can be used to graph the spectra of the (differenced) seasonally adjusted variable, the (differenced) original, and the irregular, see Figure 15.3.

A large part of the generated output can be store in the database using Test/Store in Database.
Figure 15.3  Spectral plots for ICMETI X11 decomposition
Chapter 16

Tutorial on regARIMA Modelling with X12arima for OxMetrics

16.1 Introduction

The X12arima for OxMetrics program can be used for ARIMA modelling, either with or without using the seasonal adjustment features. As an example, we first estimate the airline model, followed by a regARIMA model (ARIMA model with regressors).

16.2 Estimating the Airline model

Load the Airline.xls data set in OxMetrics (from the X12arima folder), and start X12arima (if it is not already active). Select Model/Formulate, or click on the model formulation icon on the toolbar.

The airline model is:

\[(1 - L)(1 - L^{12}) \log(y_t) = (1 - \theta_1 L)(1 - \theta_1^{12} L^{12}) \varepsilon_t.\]

In general, we write \((p \ d \ q)(P \ D \ Q)\) for the ARIMA model, so that the airline model corresponds to: \((0 \ 1 \ 1)(0 \ 1 \ 1)\).

The following steps are required to formulate the model:

1. Select the airline database, as well as the Airline variable in Formulate.

2. In the Model Settings dialog:
   - Select the log transformation.
   - Set both types of differencing to first difference. This sets \(d = D = 1\).
   - Set the ARMA Model to: Specify ARIMA model.
   - The ARMA specification is \(p = P = 0, q = Q = 1\), which is also the default.
   - Change the default number of forecasts to 24.
   - Finally, clear the X11 Seasonal adjustment check box.
3. Proceed with the default sample in Estimate.

16.2.1 Model Output

Press the Estimate button to start the iterative maximization procedure. The model output is reported in OxMetrics: the graphical analysis is in X12-Airline, as in Figure 16.1, while the estimation output is in the Results window.

MODEL DEFINITION for Airline
Transformation: Log(y)
ARIMA Model: (0 1 1)(0 1 1)

MODEL ESTIMATION/EVALUATION
Estimation converged in 6 ARMA iterations, 19 function evaluations.
ARIMA Model: (0 1 1)(0 1 1)
Nonseasonal differences: 1
Seasonal differences: 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonseasonal MA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 1</td>
<td>.4018</td>
<td>.07887</td>
</tr>
<tr>
<td>Seasonal MA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 12</td>
<td>.5569</td>
<td>.07626</td>
</tr>
<tr>
<td>Variance</td>
<td>.13481E-02</td>
<td></td>
</tr>
<tr>
<td>SE of Var</td>
<td>.16657E-03</td>
<td></td>
</tr>
</tbody>
</table>
16.2 Estimating the Airline model

Figure 16.1 Graphic analysis of airline model

Likelihood Statistics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations (nobs)</td>
<td>144</td>
</tr>
<tr>
<td>Effective number of observations (nefobs)</td>
<td>131</td>
</tr>
<tr>
<td>Number of parameters estimated (np)</td>
<td>3</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>244.6965</td>
</tr>
<tr>
<td>Transformation Adjustment</td>
<td>-735.2943</td>
</tr>
<tr>
<td>Adjusted Log likelihood (L)</td>
<td>-490.5978</td>
</tr>
<tr>
<td>AIC</td>
<td>987.1956</td>
</tr>
<tr>
<td>AICC (F-corrected-AIC)</td>
<td>987.3845</td>
</tr>
<tr>
<td>Hannan Quinn</td>
<td>990.7005</td>
</tr>
<tr>
<td>BIC</td>
<td>995.8211</td>
</tr>
</tbody>
</table>

Roots of ARIMA Model

<table>
<thead>
<tr>
<th>Root</th>
<th>Real</th>
<th>Imaginary</th>
<th>Modulus</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonseasonal MA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root 1</td>
<td>2.4888</td>
<td>0.0000</td>
<td>2.4888</td>
<td>0.0000</td>
</tr>
<tr>
<td>Seasonal MA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root 1</td>
<td>1.7955</td>
<td>0.0000</td>
<td>1.7955</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
16.3 **regARIMA Model Example**

This example uses the ‘Unit Auto Sales’ series of the U.S. Bureau of Labor Statistics. It features in Section 5.4 in the Findley, Monsell, Bell, Otto, and Chen (1998). It is example 8 (beauto0.spc) from the set of examples supplied with X-12-ARIMA by the US Census Bureau. To start, load the file `beauto.in7` as supplied with X12arima for OxMetrics.

The following steps are required to formulate the model:

1. In OxMetrics, create the following two additive outliers:
   
   ```
   "A01975.2" = dummy(1975,2,1975,2);
   "A01988.12" = dummy(1988,12,1988,12);
   ```

2. Switch to X12arima, to formulate the model with `beauto` as Y variable, and the remainder as X:

   Press OK to go to model settings

3. Select log transformation.

4. select specify ARIMA model, and formulate the airline model \((0 1 1)(0 1 1)\).

5. Under regression, add the intercept and Trading days + leap year \((td)\).

6. Also under regression, set the user type for the X variables to \(ao\).

7. As the final action in this dialog, switch X11 seasonal adjustment on:
16.3.1 Model Output

The model output is again printed in OxMetrics, and can be analyzed graphically. Figure 16.2 gives the seasonally-adjusted variable with the original in the first graph, the trend with original in the second, and the irregular in the third graph.

MODEL DEFINITION for beauto
Combined Prior Adjustment Factors
Leap Year

Transformation: Log(y)
Regression Model Constant + Trading Day + User-defined
ARIMA Model: (0 1 1)(0 1 1)

MODEL ESTIMATION/EVALUATION
Estimation converged in 42 ARMA iterations, 244 function evaluations.

Regression Model

--------------------------------------------------------------------
**Figure 16.2** Seasonal adjustment of BeAuto series

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>.0000</td>
<td>.00132</td>
<td>-.02</td>
</tr>
<tr>
<td>Trading Day</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mon</td>
<td>.0222</td>
<td>.00992</td>
<td>2.24</td>
</tr>
<tr>
<td>Tue</td>
<td>-.0125</td>
<td>.00984</td>
<td>-1.27</td>
</tr>
<tr>
<td>Wed</td>
<td>.0088</td>
<td>.01001</td>
<td>.88</td>
</tr>
<tr>
<td>Thu</td>
<td>.0313</td>
<td>.01015</td>
<td>3.09</td>
</tr>
<tr>
<td>Fri</td>
<td>-.0213</td>
<td>.00987</td>
<td>-2.15</td>
</tr>
<tr>
<td>Sat</td>
<td>-.0054</td>
<td>.00999</td>
<td>-.54</td>
</tr>
<tr>
<td>*Sun (derived)</td>
<td>-.0232</td>
<td>.00986</td>
<td>-2.36</td>
</tr>
<tr>
<td>User-defined</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User.Aug</td>
<td>-.1442</td>
<td>.04678</td>
<td>-3.08</td>
</tr>
<tr>
<td>User.Sep</td>
<td>-.3268</td>
<td>.05540</td>
<td>-5.90</td>
</tr>
<tr>
<td>User.Nov</td>
<td>.1909</td>
<td>.04907</td>
<td>3.89</td>
</tr>
<tr>
<td>User.D&amp;J</td>
<td>.2176</td>
<td>.03793</td>
<td>5.74</td>
</tr>
<tr>
<td>AO1975.2</td>
<td>.1595</td>
<td>.06123</td>
<td>2.61</td>
</tr>
<tr>
<td>AO1988.12</td>
<td>.2160</td>
<td>.06563</td>
<td>3.29</td>
</tr>
</tbody>
</table>

*For full trading-day and stable seasonal effects, the derived parameter estimate is obtained indirectly as minus the sum of the directly estimated parameters that define the effect.*

**Chi-squared Tests for Groups of Regressors**

<table>
<thead>
<tr>
<th>Regression Effect</th>
<th>df</th>
<th>Chi-Square</th>
<th>P-Value</th>
</tr>
</thead>
</table>
ARIMA Model: \((0 1 1)(0 1 1)\)
Nonseasonal differences: 1
Seasonal differences: 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonseasonal MA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 1</td>
<td>.1791</td>
<td>.06861</td>
</tr>
<tr>
<td>Seasonal MA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 12</td>
<td>.8072</td>
<td>.04473</td>
</tr>
<tr>
<td>Variance</td>
<td>.68263E-02</td>
<td></td>
</tr>
<tr>
<td>SE of Var</td>
<td>.69310E-03</td>
<td></td>
</tr>
</tbody>
</table>

Likelihood Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations (nobs)</td>
<td>207</td>
</tr>
<tr>
<td>Effective number of observations (nefobs)</td>
<td>194</td>
</tr>
<tr>
<td>Number of parameters estimated (np)</td>
<td>17</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>202.1253</td>
</tr>
<tr>
<td>Transformation Adjustment</td>
<td>-1696.4291</td>
</tr>
<tr>
<td>Adjusted Log likelihood (L)</td>
<td>-1494.3038</td>
</tr>
<tr>
<td>AIC</td>
<td>3022.6076</td>
</tr>
<tr>
<td>AICC (F-corrected-AIC)</td>
<td>3026.0848</td>
</tr>
<tr>
<td>Hannan Quinn</td>
<td>3045.1028</td>
</tr>
<tr>
<td>BIC</td>
<td>3078.1612</td>
</tr>
</tbody>
</table>

Roots of ARIMA Model

<table>
<thead>
<tr>
<th>Root</th>
<th>Real</th>
<th>Imaginary</th>
<th>Modulus</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonseasonal MA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root 1</td>
<td>5.5849</td>
<td>.0000</td>
<td>5.5849</td>
<td>.0000</td>
</tr>
<tr>
<td>Seasonal MA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root 1</td>
<td>1.2389</td>
<td>.0000</td>
<td>1.2389</td>
<td>.0000</td>
</tr>
</tbody>
</table>
References


REFERENCES


<table>
<thead>
<tr>
<th>Author</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Álvarez, J.</td>
<td>36</td>
</tr>
<tr>
<td>Abramowitz, M.</td>
<td>116</td>
</tr>
<tr>
<td>Amemiya, T.</td>
<td>7, 11</td>
</tr>
<tr>
<td>An, S.</td>
<td>119</td>
</tr>
<tr>
<td>Anderson, T. W.</td>
<td>31</td>
</tr>
<tr>
<td>Arellano, M.</td>
<td>31, 32, 34–37, 39, 47, 48, 52–54, 60, 61, 63</td>
</tr>
<tr>
<td>Baillie, R. T.</td>
<td>101</td>
</tr>
<tr>
<td>Baltagi, B. H.</td>
<td>32, 40, 45, 58</td>
</tr>
<tr>
<td>Bell, W. R.</td>
<td>130, 131, 148</td>
</tr>
<tr>
<td>Bera, A. K.</td>
<td>69</td>
</tr>
<tr>
<td>Beran, J.</td>
<td>101, 122</td>
</tr>
<tr>
<td>Bloomfield, P.</td>
<td>119</td>
</tr>
<tr>
<td>Blundell, R. W.</td>
<td>31, 34, 35, 37, 38, 54, 56, 63, 64</td>
</tr>
<tr>
<td>Bollerslev, T.</td>
<td>69, 73, 76, 82, 85, 95</td>
</tr>
<tr>
<td>Bond, S. R.</td>
<td>31, 34–39, 47, 48, 52–54, 56, 60, 63, 64</td>
</tr>
<tr>
<td>Bover, O.</td>
<td>31, 34, 37, 54</td>
</tr>
<tr>
<td>Box, G. E. P.</td>
<td>133</td>
</tr>
<tr>
<td>Brockwell, P. J.</td>
<td>102, 121</td>
</tr>
<tr>
<td>Calzolari, G.</td>
<td>96</td>
</tr>
<tr>
<td>Cao, C. Q.</td>
<td>70, 94</td>
</tr>
<tr>
<td>Chamberlain, G.</td>
<td>34</td>
</tr>
<tr>
<td>Chang, M. C.</td>
<td>102</td>
</tr>
<tr>
<td>Chen, B.-C.</td>
<td>130, 131, 148</td>
</tr>
<tr>
<td>Cheung, Y.-W.</td>
<td>120</td>
</tr>
<tr>
<td>Cox, D. R.</td>
<td>119</td>
</tr>
<tr>
<td>Cramer, J. S.</td>
<td>7, 9, 10, 15</td>
</tr>
<tr>
<td>Davis, R. A.</td>
<td>102, 121</td>
</tr>
<tr>
<td>Dickey, D. A.</td>
<td>102</td>
</tr>
<tr>
<td>Diebold, F. X.</td>
<td>120</td>
</tr>
<tr>
<td>DiNardo, J.</td>
<td>32</td>
</tr>
<tr>
<td>Doornik, J. A.</td>
<td>3, 4, 27, 77, 95, 96, 101, 103, 107, 110, 115, 118, 122, 129</td>
</tr>
<tr>
<td>Engle, R. F.</td>
<td>69, 87, 107</td>
</tr>
<tr>
<td>Findley, D. F.</td>
<td>130, 131, 148</td>
</tr>
<tr>
<td>Fiorentini, G.</td>
<td>96</td>
</tr>
<tr>
<td>Fletcher, R.</td>
<td>96</td>
</tr>
<tr>
<td>Galbraith, J. W.</td>
<td>96</td>
</tr>
<tr>
<td>Geweke, J. F.</td>
<td>117</td>
</tr>
<tr>
<td>Ghysels, E.</td>
<td>73, 76, 82</td>
</tr>
<tr>
<td>Gill, P. E.</td>
<td>96</td>
</tr>
<tr>
<td>Glosten, L. R.</td>
<td>72</td>
</tr>
<tr>
<td>Gourieroux, C.</td>
<td>69</td>
</tr>
<tr>
<td>Granger, C. W. J.</td>
<td>113, 117</td>
</tr>
<tr>
<td>Grunfeld, Y.</td>
<td>40</td>
</tr>
<tr>
<td>Hamilton, J. D.</td>
<td>69</td>
</tr>
<tr>
<td>Hansen, H.</td>
<td>77, 107, 110</td>
</tr>
<tr>
<td>Harvey, A. C.</td>
<td>129</td>
</tr>
<tr>
<td>Hausman, J. A.</td>
<td>38</td>
</tr>
<tr>
<td>He, C.</td>
<td>94</td>
</tr>
<tr>
<td>Hendry, D. F.</td>
<td>3, 4</td>
</tr>
<tr>
<td>Higgins, M. L.</td>
<td>69</td>
</tr>
<tr>
<td>Hosking, J. R. M.</td>
<td>102</td>
</tr>
<tr>
<td>Hsiao, C.</td>
<td>31, 32</td>
</tr>
<tr>
<td>Jagannathan, R.</td>
<td>72</td>
</tr>
<tr>
<td>Jenkins, G. M.</td>
<td>133</td>
</tr>
<tr>
<td>Johnston, J.</td>
<td>32</td>
</tr>
<tr>
<td>Koopman, S. J.</td>
<td>129</td>
</tr>
<tr>
<td>Ladiray, D.</td>
<td>131</td>
</tr>
<tr>
<td>Lerman, S. R.</td>
<td>11</td>
</tr>
<tr>
<td>Lilien, D. M.</td>
<td>87</td>
</tr>
<tr>
<td>Maddala, G. S.</td>
<td>11</td>
</tr>
<tr>
<td>Manski, C. F.</td>
<td>11</td>
</tr>
<tr>
<td>Maravall, A.</td>
<td>133</td>
</tr>
<tr>
<td>McFadden, D. L.</td>
<td>7</td>
</tr>
<tr>
<td>Mills, T. C.</td>
<td>69</td>
</tr>
<tr>
<td>Monsell, B. C.</td>
<td>130, 131, 148</td>
</tr>
<tr>
<td>Murray, W.</td>
<td>96</td>
</tr>
<tr>
<td>Musgrave, J. C.</td>
<td>131</td>
</tr>
<tr>
<td>Nelson, D. B.</td>
<td>69–71, 94</td>
</tr>
<tr>
<td>Newbold, P.</td>
<td>113, 117</td>
</tr>
<tr>
<td>Nickell, S. J.</td>
<td>31</td>
</tr>
<tr>
<td>Odaki, M.</td>
<td>102</td>
</tr>
<tr>
<td>Ooms, M.</td>
<td>77, 95, 96, 101, 103, 115, 118, 122, 129</td>
</tr>
</tbody>
</table>
Otto, W. R. 130, 131, 148
Panattoni, L. 96
Porter-Hudak, S. 117
Quenneville, B. 131
Reid, N. 119
Reinsel, G. C. 133
Robins, R. P. 87
Robinson, P. 101
Runkle, D. 72
Ruud, P. A. 11
Sargan, J. D. 36
Shephard, N. 69, 129
Shiskin, J. 131
Sowell, F. 116
Srba, F. 3
Stegun, I. A. 116
Teräsvirta, T. 94
Tsay, R. S. 69
Vuong, Q. H. 11
White, H. 35, 61
Windmeijer, F. 53, 56, 62
Wooldridge, J. M. 95
Wright, M. H. 96
Young, A. H. 131
Zinde-Walsh, V. 96
Subject Index

abdata.in7 47
Airline model 105, 133, 145
Akaike information criterion (AIC) 76, 120
ARCH 69, 107
ARFIMA 101–125
— Z variables 121
— model 115
Autocovariance function 102, 116
Covariance stationarity 102
Estimation output 120
Exact maximum likelihood 118
Fixing d 105
Fixing AR lags 105
Fixing parameters 121
Forecasting 122
— Levels — 122
— Naive — 122
Initial values 117
Long memory 102
Modified profile likelihood 119
Non-linear least squares 119
Portmanteau test 107
Regressors in mean 117
Sample mean versus known mean 120
Specification of order 105
Treatment of mean 105, 120
Variance-covariance matrix 119
Weighted estimation 121
Arfima
— Estimation options 120
ARMA
— model 101, 115, 132
— model selection 133
Autocorrelation function 116
Autocovariance function 115
Invertibility 102
Stationarity 102
Automatic model selection 27
Autometrics 27
Batch 64, 97, 123
— Variable names 66, 98, 124
Batch (ARFIMA) 123
Batch (GARCH) 97
Batch (Panel models) 64
BFGS 96, 118
Box–Jenkins method 101, 133
Categorical dependent variable 7
Discrete choice
— model 7
— Binary — 7
— Multinomial — 8
EGARCH 71
— in mean 71
— with GED errors 71, 86, 94
exch_debp.in7 73
Fractional difference operator 102
GARCH 69–98
— covariance 95
— in mean 71, 87
— parameter restrictions 94
— with Student-t errors 85, 94
ARMA representation 70
Asymmetric — 72, 90
Forecasting 70, 84
Impose alpha,beta >= 0 95
Impose conditional variance >= 0 94
Impose stationarity and alpha+beta >= 0 95
Non-normal error distribution 85
Portmanteau test 78
Start-up of variance recursion 95
Threshold — 72, 90
Unrestricted estimation 80, 94
General-to-specific (Gets) 27
Generalized error distribution 94
grunfeld.xls 40, 42
Hessian matrix 95
HMSE 76
### SUBJECT INDEX

<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypergeometric function</td>
<td>116</td>
</tr>
<tr>
<td>I(d)</td>
<td>102</td>
</tr>
<tr>
<td>ICMETI.in7</td>
<td>136</td>
</tr>
<tr>
<td>IGARCH</td>
<td>71, 76, 81</td>
</tr>
<tr>
<td>Information matrix</td>
<td>95</td>
</tr>
<tr>
<td>Instrument</td>
<td>49</td>
</tr>
<tr>
<td>Level</td>
<td>49, 53</td>
</tr>
<tr>
<td>Logit</td>
<td>7–28</td>
</tr>
<tr>
<td>— estimated probabilities</td>
<td>11</td>
</tr>
<tr>
<td>— likelihood ratio tests</td>
<td>11</td>
</tr>
<tr>
<td>— weighted estimation</td>
<td>10</td>
</tr>
<tr>
<td>Alternative dependent regressor</td>
<td>9, 11, 26</td>
</tr>
<tr>
<td>Alternative independent regressor</td>
<td>9, 25</td>
</tr>
<tr>
<td>Alternative specific coefficient</td>
<td>9</td>
</tr>
<tr>
<td>Automatic model selection</td>
<td>27</td>
</tr>
<tr>
<td>Baseline model</td>
<td>11</td>
</tr>
<tr>
<td>Binary — model</td>
<td>8, 17</td>
</tr>
<tr>
<td>Choice-based sampling</td>
<td>11</td>
</tr>
<tr>
<td>Conditional — model</td>
<td>10, 24</td>
</tr>
<tr>
<td>Derivatives of probabilities</td>
<td>12</td>
</tr>
<tr>
<td>Elasticities</td>
<td>13</td>
</tr>
<tr>
<td>Generic coefficient</td>
<td>9</td>
</tr>
<tr>
<td>Grouped — model</td>
<td>12, 16, 21</td>
</tr>
<tr>
<td>IIA property</td>
<td>10</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>8, 9</td>
</tr>
<tr>
<td>Log-odds ratio</td>
<td>8, 10</td>
</tr>
<tr>
<td>Multinomial — model</td>
<td>10, 22</td>
</tr>
<tr>
<td>Newton’s method</td>
<td>18</td>
</tr>
<tr>
<td>Norm observations</td>
<td>13</td>
</tr>
<tr>
<td>Normalization</td>
<td>10</td>
</tr>
<tr>
<td>Observed versus predicted</td>
<td>14</td>
</tr>
<tr>
<td>Outlier analysis</td>
<td>14</td>
</tr>
<tr>
<td>Quasi-elasticities</td>
<td>13</td>
</tr>
<tr>
<td>Reference state</td>
<td>10</td>
</tr>
<tr>
<td>Saturated model</td>
<td>12</td>
</tr>
<tr>
<td>Zeroline model</td>
<td>11</td>
</tr>
<tr>
<td>Long memory</td>
<td>102</td>
</tr>
<tr>
<td>Moving average</td>
<td>112</td>
</tr>
<tr>
<td>NLCAR.in7</td>
<td>15, 17</td>
</tr>
<tr>
<td>NLCARGROUPED.in7</td>
<td>21</td>
</tr>
<tr>
<td>Outer product of gradients</td>
<td>95</td>
</tr>
<tr>
<td>Overdifferencing</td>
<td>102</td>
</tr>
<tr>
<td>Ox</td>
<td>4, 101</td>
</tr>
<tr>
<td>OxMetrics</td>
<td>3</td>
</tr>
<tr>
<td>OxMetrics dialogs</td>
<td>92</td>
</tr>
<tr>
<td>Generate Ox Code</td>
<td>93</td>
</tr>
<tr>
<td>Model menu</td>
<td>92</td>
</tr>
<tr>
<td>Batch</td>
<td>92</td>
</tr>
<tr>
<td>Ox Batch Code</td>
<td>93</td>
</tr>
<tr>
<td>Panel data</td>
<td>31–66</td>
</tr>
<tr>
<td>— combined estimation</td>
<td>55, 63</td>
</tr>
<tr>
<td>Balanced</td>
<td>31</td>
</tr>
<tr>
<td>Between-groups estimation</td>
<td>32, 44, 58, 59</td>
</tr>
<tr>
<td>Common factor</td>
<td>33</td>
</tr>
<tr>
<td>Data organization</td>
<td>41</td>
</tr>
<tr>
<td>Dynamic — estimators</td>
<td>33, 59</td>
</tr>
<tr>
<td>GLS estimation</td>
<td>32, 58</td>
</tr>
<tr>
<td>GMM estimation</td>
<td>33, 34</td>
</tr>
<tr>
<td>1-step</td>
<td>35, 38, 48, 61</td>
</tr>
<tr>
<td>2-step</td>
<td>35, 38, 52, 61</td>
</tr>
<tr>
<td>Combined</td>
<td>37, 54, 55, 63</td>
</tr>
<tr>
<td>Standard errors</td>
<td>35, 61</td>
</tr>
<tr>
<td>Heteroscedasticity</td>
<td>33</td>
</tr>
<tr>
<td>Instrumental variable</td>
<td>33</td>
</tr>
<tr>
<td>IV estimation</td>
<td>53</td>
</tr>
<tr>
<td>LSDV estimation</td>
<td>32, 44, 58</td>
</tr>
<tr>
<td>ML estimation</td>
<td>32, 59</td>
</tr>
<tr>
<td>OLS estimation</td>
<td>32, 58</td>
</tr>
<tr>
<td>OLS on differences estimation</td>
<td>44</td>
</tr>
<tr>
<td>Pooled estimation</td>
<td>44</td>
</tr>
<tr>
<td>Static — estimators</td>
<td>32, 58</td>
</tr>
<tr>
<td>Test</td>
<td>36</td>
</tr>
<tr>
<td>— for overidentifying restrictions</td>
<td>36</td>
</tr>
<tr>
<td>— of significance</td>
<td>63</td>
</tr>
<tr>
<td>AR</td>
<td>44, 62, 64</td>
</tr>
<tr>
<td>Sargan</td>
<td>36, 38, 63</td>
</tr>
<tr>
<td>Wald (dummy)</td>
<td>52, 63</td>
</tr>
<tr>
<td>Wald (joint)</td>
<td>52, 63</td>
</tr>
<tr>
<td>Wald (time)</td>
<td>52, 63</td>
</tr>
<tr>
<td>Transformations</td>
<td>57</td>
</tr>
<tr>
<td>Deviation from time means</td>
<td>57</td>
</tr>
<tr>
<td>First differences</td>
<td>35, 57</td>
</tr>
<tr>
<td>GLS deviations</td>
<td>57</td>
</tr>
<tr>
<td>Orthogonal deviations</td>
<td>34, 35, 57</td>
</tr>
<tr>
<td>Unbalanced</td>
<td>31</td>
</tr>
<tr>
<td>Within-groups estimation</td>
<td>32, 36, 44, 58, 59</td>
</tr>
<tr>
<td>PcGive dialogs</td>
<td>18, 43, 51, 75, 79, 106</td>
</tr>
<tr>
<td>Estimate Model</td>
<td>18, 43, 51, 75, 79, 106</td>
</tr>
<tr>
<td>Forecast</td>
<td>84, 111, 113</td>
</tr>
<tr>
<td>Formulate Model</td>
<td>18, 21, 26, 43, 49, 53, 75, 82, 105</td>
</tr>
<tr>
<td>Functions</td>
<td>50, 54, 55</td>
</tr>
<tr>
<td>Further Output</td>
<td>20</td>
</tr>
<tr>
<td>Graphic Analysis</td>
<td>24, 77, 109</td>
</tr>
<tr>
<td>Linear Restrictions</td>
<td>108</td>
</tr>
<tr>
<td>Model Options</td>
<td>19</td>
</tr>
</tbody>
</table>
Model Settings 18, 43, 51, 75, 81, 85, 87, 105, 106
Norm Observation 23
Recursive Graphics 79
Store in Database 112
Test 78, 110
PcGive menus
Model menu
  Dynamic Panel Methods 49
  Estimate 18, 43, 51, 75, 106
  Formulate 18, 21, 26, 43, 49, 53, 75, 82, 105
Model Settings 18, 43, 51, 75, 81, 85, 87, 94, 105, 106
Options 19
Static Panel Methods 42
Package menu
  Volatility Models (Garch) 74
Test menu
  Forecast 84, 111, 113
  Further Output 20
  Graphic Analysis 24, 77, 109
  Linear Restrictions 108
  Norm Observation 23
  Recursive Graphics 79
  Store in Database 112
  Test 78, 110
  Test Summary 77, 107
Pochhammer’s symbol 116
Probit
  Binary — model 8, 20
  Derivatives of probabilities 12
  Grouped — model 12, 16
Quick Seasonal Adjustment 136
regARIMA 132
Residual
  Store — in database 65, 97, 124
RSS 61, 64
Schwarz criterion 27
Seasonality 73
STAMP 129
Standard error
  Robust — 76
t-distribution 94
Test
  Dickey–Fuller — 101
  Instability — 101
  Normality — 77, 107
  Portmanteau 78, 107
  Wald (dummy) 44
  Wald (joint) 44
  Thick tails 71, 79
  Toeplitz matrix 116
  TRAMO 129, 133
  TSS 61, 64
  UKrpiq.in7 103
  USyieldm.in7 87
  Utility model 9
  Variable names in batch 66, 98, 124
  Volatility clustering 73
X-11 129, 131
X-11-ARIMA 129
X-12-ARIMA 129, 132
X12arima 105, 129–151
  Credits 130
  Disclaimer 130
  Documentation 130
  Limitations 130
  Maximum sample size 130
  Model Menu
    Model Settings 146
  Quick Seasonal Adjustment 136
  Test Menu 143