Multivariate time series

Adam Boult (www.bou.lt)

April 29, 2024

Contents

Part I

Multivariate stochastic processes

Chapter 1

Multivariate time series

1.1 Multiple time series

1.1.1 Cointegration

If we have multiple variables, we can explore the order of integration of linear combinations.

If two series have time trends, a linear combination of them could remove this.

1.1.2 Exogeneity

Contemporaneous exogeneity

$Cov(x_{it}, u_{it}) = 0$

Strict exogeneity

$Cov(x_{is}, u_{it}) = 0)$

This is stronger than contemporeous, all periods.

Shocks don't affect future outcomes.

Sequential exogeneity

Sequential exogeneity: a bit looser than strict exogeneity. only holds when $s \leq t$.

So shocks can affect, but only in future.

1.1.3 Introduction

Weak stationary processes can be decomposed to a deterministic and a stochastic component.

Chapter 2

Bayesian networks

2.1 Bayesian networks

2.1.1 Bayesian networks

Part II

Multivariate discrete-time stochastic processes

Chapter 3

Vector Autoregression (VAR)

3.1 Vector Autoregression (VAR)

3.1.1 Vector Autoregression (VAR)

We consider a vector of observables, not just one

Autoregressive (AR) model for a vector.

VAR(p) looks p back.

The AR(p) model is:

$y_t = \alpha + \sum_{i=1}^{p} \beta y_{t-i} + \epsilon_t$

VAR(p) generalises this to where y_t is a vector. We define VAR(p) as:

y_t

$y_t = c + \sum_{i=1}^{p} A_i y_{t-i} + \epsilon_t$

3.1.2 VAR impulse response

3.1.3 Bayesian VAR

3.2 Structural models

3.2.1 Autoregressive Distributed Lag (ARDL) model

Include lagged y and lagged x (and current x)

Chapter 4

ARMAX

4.1 ARMAX

4.1.1 ARMAX

4.1.2 ARIMAX

4.1.3 SARIMA

Chapter 5

Partial Adjustment Model (PAM)

5.1 Partial Adjustment Model

5.1.1 Partial Adjustment Model

Estimating a static model

We start by estimating a static model.

$y_t = \alpha + \theta x_t + \gamma_t$

Equilibrium

We then use this form an equilibrium for y_t, y_t^*.

$y_t^* = \hat{\alpha} + \hat{\theta} x_t$

The process depends on the difference from this equilibrium.

$y_t - y_{t-1} = \beta(y_t^* - y_{t-1}) + \epsilon_t$

$y_t - y_{t-1} = \beta(\hat{\alpha} + \hat{\theta} x_t - y_{t-1}) + \epsilon_t$

$y_t = \beta\hat{\alpha} + \beta\hat{\theta} x_t + (1-\beta) y_{t-1} + \epsilon_t$

$y_t = \alpha y_{t-1} + (1-\beta)(y_t^* - y_{t-1}) + \epsilon$

The higher β, the slower the adjustment.

If stationary, can we can use OLS.

Chapter 6

Error Correction Model

6.1 Error Correction Model

6.1.1 Error Correction Model

Static model

Like PAM we start with static estimator.

The ECM

The ECM does a regression with first differences, and includes lagged error terms.

We start with a basic first-difference model.

$\Delta y_t = \Delta x_t$

We could also expand this to include laggs for both x and y. Here we don't.

We know that long term $y_t = \theta x_t$. We use the error from this in a first difference model.

$\Delta y_t = \alpha \Delta x_t + \beta(y_{t-1} - \theta x_{t-1})$

Page on identifying error terms

Also, page on Vector Error Correction Model (VECM)

Part III

Estimating multivariate time series models

Chapter 7

Multivariate forecasting

7.1 Introduction to multiple time series

7.1.1 Testing for cointegration with Johansen

7.2 Vector Autoregression (VAR)

7.2.1 Vector Autoregression (VAR)

We consider a vector of observables, not just one

Autoregressive (AR) model for a vector.

VAR(p) looks p back.

The AR(p) model is:

$y_t = \alpha + \sum_{i=1}^{p} \beta y_{t-i} + \epsilon_t$

VAR(p) generalises this to where y_t is a vector. We define VAR(p) as:

y_t

$y_t = c + \sum_{i=1}^{p} A_i y_{t-i} + \epsilon_t$

7.2.2 VAR impulse response

7.2.3 Bayesian VAR

7.3 Structural models

7.3.1 Autoregressive Distributed Lag (ARDL) model

Include lagged y and lagged x (and current x)

If the processes are stationary, then we can use OLS. THIS IS A BROADER POINT! INTRO??

7.4 ARMAX

7.4.1 ARMAX

7.4.2 Error Correction Model

Static model

Like PAM we start with static estimator.

The ECM

The ECM does a regression with first differences, and includes lagged error terms.

We start with a basic first-difference model.

$\Delta y_t = \Delta x_t$

We could also expand this to include laggs for both x and y. Here we don't.

We know that long term $y_t = \theta x_t$. We use the error from this in a first difference model.

$\Delta y_t = \alpha \Delta x_t + \beta(y_{t-1} - \theta x_{t-1})$

Page on identifying error terms

Also, page on Vector Error Correction Model (VECM)

7.4.3 Partial Adjustment Model

Estimating a static model

We start by estimating a static model.

$y_t = \alpha + \theta x_t + \gamma_t$

Equilibrium

We then use this form an equilibrium for y_t, y_t^*.

$y_t^* = \hat{\alpha} + \hat{\theta} x_t$

The process depends on the difference from this equilibrium.

$y_t - y_{t-1} = \beta(y_t^* - y_{t-1}) + \epsilon_t$

$y_t - y_{t-1} = \beta(\hat{\alpha} + \hat{\theta} x_t - y_{t-1}) + \epsilon_t$

$y_t = \beta\hat{\alpha} + \beta\hat{\theta}x_t + (1-\beta)y_{t-1} + \epsilon_t$

$y_t = \alpha y_{t-1} + (1-\beta)(y_t^* - y_{t-1}) + \epsilon$

The higher β, the slower the adjustment.

If stationary, can we can use OLS.

Part IV

Advanced inference (time)

Chapter 8

Homogeneous treatment effects

8.1 Introduction

8.1.1 Treatment data

Recap

With multilevel data with fixed coefficients we have:

$y_{ij} = \mathbf{x}_{ij}\theta + m_j + \epsilon_{ij}$

We can estimate m_j using fixed effects or similar methods.

Treatment data

If the data is grouped by whether an entity was treated then will have:

- y_{i0} - the outcome if the entity was not treated
- y_{i1} - the outcome if the entity was treated

However we only observe y_i and D_i.

$y_i = y_{i0} + D_i(y_{i1} - y_{10})$

8.1.2 Average Treatment Effects (ATE, ATET, ATEUT)

Average Treatment Effect (ATE)

$ATE = E[y_{i1} - y_{i0}]$

Average Treatment Effect on the Treated (ATET)

$ATE = E[y_{i1} - y_{i0}|D_i = 1]$

$ATE = E[y_{i1}|D_i = 1] - E[y_{i0}|D_i = 1]$

Average Treatment Effect on the Untreated (ATEUT)

8.1.3 Conditional Average Treatment Effect (CATE)

$E[y_{i1} - y_{i0}|\mathbf{x}_i]$

8.2 Exogenous treatment

8.2.1 Randomly Controlled Trials (RCTs)

If the model is:

$y_i = D_i\theta + g(X) + \epsilon_i$

And D is randomly assigned, then we can estimate

$y_i = D_i\theta + \epsilon_i$

To get an estimate for θ without collecting data on X.

8.2.2 Calculating CATEs in RCTs with interaction terms

8.2.3 Calculating CATEs in RCTs with subgroup analysis

8.3 Calculating treatment effects without estimating missing data

8.3.1 Regression

We can simply regress outcomes on variables, including treatment.

This assumes treatment effects are constant.

This also assumes that outcomes y_{1i} and y_{0i} are independent of D_i, conditional on X.

If we are missing variables in X then we will have biased estimates.

This also assumes the effects of X are linear.

We assume: $E[y_{0i}|\mathbf{x}_i, D_i] = \mathbf{x}_i\theta$.

8.3.2 Instrumental Variables and natural experiments

8.3.3 Regression discontinuity

8.3.4 Synthetic controls

8.4 Calculating treatment effects by estimating missing data

8.4.1 Matching

Matching is similar to regression. We assume that effects are constant, and the effect of treatment on y_{0i} and y_{1i} are independent of treatment, once controlling for X.

Again, this is biased if this is not the case.

We however do not have to assume a linear form for X.

We assume: $E[y_{ji}|\mathbf{x}_i, D_i] = E[y_{ji}|\mathbf{x}_i]$

For each entity, find a near entity which had the opposite treatment.

8.4.2 Propensity score matching

Match on the chance of getting treatment, given covariates.

8.4.3 Matrix completion

$E[y_{i1} - y_{i0}|\mathbf{x}_i]$

8.5 Using semi-parametric

8.6 Other

8.6.1 Estimating ATE using MCMC

8.6.2 Local Average Treatment Effect (LATE)

We have IVs for treatment.

8.6.3 Treatment effects

+ propsentiy score weighting + regression adjustemnt + matching + IV + Regression discontinuity

8.6.4 Meta analysis

big page in advanced analytics? Random effects meta analysis?

meta analysis: fixed effect v random effects model

types of study: + RCT + cohort studies + case-control studies + cross sectional studies

8.6.5 Dose response curve

8.6.6 Sensitivity analysis

8.6.7 Page on Rubin causal model

Chapter 9

Heterogeneous treatment effects

9.1 Heterogenous treatment effects

9.1.1 Introduction

9.1.2 subgroup analysis

9.1.3 interaction terms

9.1.4 efficient policy learning

9.1.5 Het DML

$y = a(z) + db(z)$ Het effects is $b(z)$ We build groups instead of arbitrary function. So we estimate $E[b(z)|G]$

Use part of the data set to estimate

$\hat{y} = \hat{a}(z) + D\hat{b}(z)$

Use $s = \hat{b}(z)$ to stratify. Key point is defining subgroups algorithmically. Less opportunity for hacking

9.1.6 Continuous treatment effects

9.1.7 Intent-to-treat

9.2 (LATE, causal tree (from CART))

9.2.1 Introduction

bart causal is different to causal tree

In stuff now two problems: + non random but constant effect + Random but heterogenous effect

causal trees can find heterogenous treatment effects

Approaches: We have treated and untreated. X and y Estimate y—x for treated, and untreated separately. Then take differnece for a given x to be the estimated treatment effect

2nd approach: have treatment as input diffence is again y—x - y—x treatment minus no treatment

3rd approach: (type of single tree) split not by predictive power, but by treatment effect differnece

4th approach: cross validation at each leaft we note the sample average treatment effect goal is to choose hyper parameters which minimise sum of diffence between these and cross valid data

Once we have the trees from the last one, calculate the effect using test data. nb: separate creating of tree to estimation of treatment effect

9.2.2 Instrumental forests

Estimate LATE

like causal forest, but do IV regression on leaf.

Chapter 10

Causal trees

10.1 Causal trees

10.1.1 Measuring treatment effects in leaves

10.1.2 Sample splitting for treatment effects

10.1.3 Honest trees

We use part of the sample to estimate Θ, and another part of the sample to estimate the treatment effect.

10.1.4 Estimating ATE using MCMC

10.2 Ensemble methods for causal trees

10.2.1 Causal forests

10.2.2 Bayesian causal forests

Part V

Other

Chapter 11

Weather forecasting

11.1 Weather