Regression Analysis III: Advanced Methods
Syllabus - Summer 2010

Instructor:

Dave Armstrong
Department of Political Science
University of Wisconsin–Milwaukee

e: armstrod@uwm.edu
Course website: http://www.quantoid.net/ICPSR.php

Teaching Assistant:

Matthew Painter
Department of Sociology
(The) Ohio State University

e: painter.63@sociology.osu.edu

1 Overview and Course Objectives

The Regression III course takes a considerably different form than the first two regression courses at the Summer Program. This course will hopefully prepare you for the things you will encounter when you (attempt to) publish quantitative work with linear models. Initial linear model classes focus on the assumptions and theoretical considerations of linear models and generally walk you through estimation and interpretation. Good courses also deal with diagnostics, though these often get less time than they should. Further, it is not always obvious what violations of these assumptions will lead to in practical terms. This course will provide you with a systematic approach to assessing, fixing and presenting your linear model results. Though we focus almost exclusively on the linear model (we will allude to nonlinear models occasionally), the logic we follow will be helpful in dealing with nonlinear models as well.

This is a class that deals exclusively with observational data - those not collected in experimentally controlled environments. As such, we will spend little time on ANOVA and no time at all talking about concerns that are specific to the analysis of experimental data.

2 Requirements

This course is a practical, data-analytic extension of what you learned in your department's linear models class or the Regression II class at the ICPSR Summer Program. As such, I assume you are familiar with the types of things taught in these courses - Gauss-Markov assumptions, properties of OLS estimators, and statistical inference for linear model coefficients. While I assume this knowledge exists, I will spend time reviewing these results briefly in class. If you are not sure where you belong in the series of linear models courses at the Summer Program, please se me or the Summer Program director and we will make sure you end up the most appropriate class.

Participants being formally assessed in this course course will the **R** statistical computing environment exclusively. This is not optional. This implies two things:

1. 1.We will spend time going over R-code to generate the results we talk about in class. The class will be theoretically driven, but will have a considerable amount of associated computer work. If you are looking for a purely theoretical class, this is very likely not the one for you.

2. It is essential for you to know or be learning the **R** language. Many of the results we generate in class are either impossible or (more likely) considerably more difficult to generate in many other general purpose statistical software packages (e.g., Stata and SPSS). John Fox will be teaching an Introduction to **R** from the 21^{st} of June to the 2^{nd} of July. John is probably one of the smartest **R** people you will ever meet, so I would encourage you in the strongest possible way to take John's course, even if you have some background in R already.

Obviously, those not being formally assessed in the course can attend without making a commitment to learn **R** . I have no problem with this, but please be aware that this will have a considerable (negative) impact on your ability to employ some of these techniques when you go back to your home institution. We will have lab classes that will allow you to get some hands-on practice with the results we generate in class.

If you're one of those "glutton-for-punishment" types, you may also find it useful to learn LaTeX. LaTeXis a system for typesetting documents. People find it most useful for typesetting documents that are heavy on mathematical notation, but this is just the tip of the iceberg. LaTeXhas its own bibliographic software (BibTeX) and will automatically build (and re-build) tables of contents, lists of figures and lists of tables. It also automatically numbers (and re-numbers when necessary) tables, figures and equations, changing appropriately formed references to those objects when table, figure or equation numbers change. Best of all, the main LaTeXtypesetting engine is free as are a number of powerful text editors that serve as front-ends to this software. Everything I present in class is written in LaTeX; specifically, the slides are all made with a package called "Beamer". Further, there are those who see LaTeXas a sort of secret handshake for nerds. So, if you want to be one of the "cool" kids, then you should definitely try it; everyone else is doing it.

3 Course Text(s)

No one text effectively presents all of the material that will be covered in this course. That said, much of the material is covered in:

Fox, John. (2008) Applied Regression Analysis and Generalized Linear Model. *2nd ed.* Thousand Oaks, CA: Sage Publications, Inc.

Fox, John. (2002) An R and S-PLUS Companion to Applied Regression. Thousand Oaks, CA: Sage Publications, Inc.

The R and S-PLUS Companion is a great book for those currently learning **R**. This is widely recognized as one of the best ways for Social Scientists to get into **R**. The Applied Regression book is a great general purpose regression book. Much of what we talk about will be covered in other regression books. If you've got a particular favorite, then it might be worth supplementing your reading from your chosen regression book with pieces from the Fox book that are not covered by your favorite. Some books that I think are pretty good (depending on your orientation toward visualization, etc...) are:

Gujarati, Damodar N. (2002) Basic Econometrics. *4th ed.* New York: McGraw Hill/Irwin.

Wooldridge, Jeffrey M. (2005) Introductory Econometrics. *3rd ed.* Mason, OH: Southwestern.

Cook, R. Dennis and Sanford Weisberg. (1999) Applied Regression Including Computing and Graphics. New York: Wiley & Sons, Inc.

We will also use a number of other books and articles to deal with more specialized issues. These are listed below (along with the appropriate chapters/pages) for the classes in which we use them.

4 Software

One of **R**'s main virtues from the grad-student point of view is that the base package and all of the add-ons (called libraries in **R**) are free. You can download the base package of **R** from the Comprehensive R Archive Network (CRAN) website `<http://www.cran.r-project.org>`. As of this writing, the most recent version is 2.7.0. **R** is updated a couple of times per year so you'll have to look back here periodically for updates. We will be using a number of add-on libraries, including:

`boot` Bootstrapping techniques associated with Davison and Hinkley (1997).

`car` Regression diagnostics and other procedures associated with Fox (2002).

`effects` John Fox's GLM effects library.

`foreign` Imports data from SPSS, Stata and SAS.

`leaps` Methods for subset selection following Miller (2002).

`MASS` Modern statistical procedures from Venables and Ripley (2002).

`mgcv` Methods for Generalized Additive Models following Wood (2006).

`nlme` **and** `lme4` Methods for linear and non-linear mixed effects models following Pinheiro and Bates (2000).

`qvcalc` Package to calculate quasi-variances from Firth (2003), Firth and Menzes (2004).

`Rcmdr` John Fox's GUI front-end for **R**.

`relimp` Package written by David Firth to implement procedures in Silber, Rosenbaum and Ross (1995).

`robustbase` Package to implement robust regression, as in Yohai (1987).

`rpanel` User-friendly front-end to TCL/TK widgets.

`RWinEdt` Package that creates links between **R** and WinEDT.

`scatterplot3d` Tools for creating 3-dimensional scatterplots.

`splines` Package for creating smoothing splines.

Many of the above packages are included with the base distribution of **R**. However, some must be downloaded separately from CRAN. The easiest way to do this is first install the base package and then use the menus within **R** to automatically download and install required packages:

`Packages` → `Install package(s) from CRAN...`

4.1 Related Software

A good text editor is invaluable when using **R** and LaTeX. I use WinEDT (on Windows) and TextMate (on Mac) for both of these tasks. Both of these come with trial periods to allow you to test them out and neither is prohibitively expensive. WinEDT can be downloaded from `<http://www.winedt.com/>` and TextMate can be downloaded from `<http://macromates.com/>`. Emacs is another useful editor that can be used for both LaTeXand **R**. You can get information on how to install the **R** implementation of this (Emacs Speaks Statistics), from `<http://ess.r-project.org/>`. If you follow the directions, you should be able to get this up and running in relatively little time. There are other editors and you're certainly welcomed to use any other one you want, but these are the ones I know.

5 Course Schedule

Each entry represents a single day in the schedule. Readings are designated either as suggested ($*$) or supplemental ($-$). For most of you, this is not the only class you are taking and as the weeks fly by, your time will undoubtedly be too limited to read everything indicated in the syllabus. However, this should serve as a nice reference to which you can return if the intricacies of a particular topic have faded from your memory.

1. Preliminary Material (Tuesday, June 22)

   (a) Goals for the course

   (b) Getting started with **R**.

   Readings:

   $*$ Fox (2008), Chapters 1 & 2
   $*$ Fox (2002), Chapters 1 & 2
   $-$ Venables and Ripley (2002), Chapters 1-3

2. OLS I: The Basics of Least Squares Regression (Wednesday, June 23)

   (a) Least-squares fit

   (b) Properties of the least-squares estimator

   (c) Statistical inference

   (d) Regression in matrix form

   Readings:

   $*$ Clarke (2005)
   $*$ Fox (2008), Chapters 5, 6 & 9
   $*$ Fox (2002), Chapter 4
   $*$ Gill (1999)
   $-$ Gelman and Stern (2006)
   $-$ Abbott (1988, 1998)
   $-$ Lewis-Beck and Skalaban (1990), Achen (1990), King (1990)

3. Graphics (Thursday, June 24)

   (a) Traditional, Lattice and Grid graphics

   (b) Types of **R** graphs

   (c) Graphical elements

   (d) Building **R** graphs

   Readings:

   ∗ Fox (2002) Chapter 7
   ∗ Jacoby (1997, 1998, 2006)
   ∗ Murrell (2006) Chapters 1-4
   – Kastellec and Leoni (2007)
   – Venables and Ripley (2002) Chapter 4

4. OLS II: Effective Presentation (Friday, June 25)

   (a) Factors and contrsts; quasi-variances and graphical displays for dummy regressors

   (b) Fitted values, interactions and effect displays

   (c) Standardization and relative importance

   Readings:

   ∗ Brambor, Clark and Golder (2006)
   ∗ Firth (2003)
   ∗ Silber, Rosenbaum and Ross (1995) – Braumoeller (2004)
   – Firth and Menzes (2004)
   – Kam and Franzese (2007)

5. Diagnostics I: Linearity (Monday, June 28)

   (a) Diagnosing linearity through residual plots

   (b) Fixing non-linearity with data transformations

   (c) Linearity and ordinal variables

   Readings:

   * Fox (2008) Chapters 4 & 12 (Sections 12.3-12.5)
   * Fox (2002) Chapter 3
   * Jacoby (1999)
   – Cook and Weisberg (1999) Chapter 16
   – Box and Tidwell (1962)

6. Non-Linearity, Smoothing and Splines (Tuesday, June 29)

   (a) Nonparametric Smoothing - Lowess

   (b) Inference for regression smoothers

   (c) Regression Splines

   Readings:

   * Fox (2008) Chapters 17 & 18
   * Keele (2008) Chapters 2 & 3
   – Fox (2000*b*,*a*)

7. Generalized Additive Models (Wednesday, June 30)

   (a) Estimation and Backfitting

   (b) Degrees of freedom

   (c) Cross-validation for smoothing parameters

   (d) Diagnostics

   Readings:

   * Fox (2000*a*)
   * Keele (2008) Chapters 4-6
   – Wood (2006)
   – Hastie and Tibshirani (1990)

8. Lab I: (Thursday, July 1)

   (a) Non-linearity transformations

   (b) Smoothers and splines

   (c) Generalized additive models

9. Diagnostics II: Outliers and Influential Data (Friday, July 2)

   (a) Outliers, leverage and influential data

   (b) Hat values, standardized residuals, Cook's D

   Readings:

   ∗ Fox (2008) Chapter 11
   ∗ Fox (2002) Chapter 6 (pp 101-201)
   ∗ Cook and Weisberg (1999) Chapter 15
   – Jasso (1985, 1996), Kahn and Udry (1986)

10. Fixing Outliers and Influential Data: Robust Regression (Tuesday, July 6)

    (a) Breakdown point, influence function and various types of robust regression

    (b) M-estimation (and extension) and interative reweighted least squares

    (c) Diagnostics for outliers revisited

    (d) Robust GLMs

    Readings:

    ∗ Andersen (2008)
    ∗ Fox (2008) Chapter 19
    – Cantoni and Ronchetti (2001)
    – Rousseeuw and Leroy (1987)

11. Diagnostics III: Non-constant error variance and collinearity (Wednesday, July 7)

    (a) Residual plots

    (b) ML transformations of Y

    (c) Weighted least squares

    (d) Heteroskedastic linear regression

    (e) Robust standard errors

    Readings:

    * Fox (2008) Chapters 12 & 13
    * Fox (2002) Chapters 3 & 6
    * Long and Ervin (2000)
    * Harvey (1976)
    – Cook and Weisberg (1999) Chapter 14

12. Model Selection (Thursday, July 8)

    (a) Theoretical issues in model searching and post-data model construction

    (b) Model selection criteria and multi-model inference.

    (c) Subset selection models

    Readings:

    * Box (1976)
    * Burnham and Anderson (2004)
    * Fox (2008) Chapter 22
    * Miller (2002)
    * Leamer (1974, 1978, 1983), Leamer and Leonard (1983)
    – Mallows (1974), Miller (1984)
    – Freedman (1991*b*,*a*), Berk (1991), Blalock (1991), Mason (1991)

13. Lab II (Friday, July 9)

    (a) Outliers and Robust Regression

    (b) Heteroskedasticity

    (c) Model Selection

14. Handling Dependent Data (Monday, July 12)

    (a) Mixed-effects models for clustered and longitudinal data

    (b) Robust standard errors revisited

    Readings:

    ∗ Snijders and Bosker (1999) Chapters 1-5
    ∗ Gelman and Hill (2007) Chapters 11-13
    – Raudenbush and Bryk (2002) Chapters 1,2, & 4
    – Pinheiro and Bates (2000) Chapters 1 & 2

15. Multilevel Models for Longitudinal Data (Tuesday, July 13)

    Readings

    ∗ Raudenbush and Bryk (2002) Chapter 6
    ∗ Singer and Willett (2003) Part I (especially Chapters 1-5)
    ∗ Bliese and Ployhart (2002)
    – Barnes et al. (2006)
    – Downey, vonHipple and Beckett (2004)
    – Curran (2003)
    – Kim and Sakamoto (2008)

16. Fixed- and Random-effects Models for TSCS Data (Wednesday, July 14)

    (a) Fixed Effects assumptions

    (b) Random effects assumptions

    (c) Choosing between fixed- and random-effects

    Readings

    ∗ Beck and Katz (1995) ∗ Beck and Katz (2007) ∗ Plumper and Troeger (2007) ∗ Wilson and Butler (2007) ∗ Wooldridge (2002) Chapters 10 & 11
    – Hsiao (1986)

17. Missing Data and Multiple Imputation (Thursday, July 15)

    (a) What's the problem with missing data?

    (b) When can we fix it?

    (c) How do we impute the data and use those imputations?

    Readings:

    ∗ Allison (2001)
    ∗ Schafer (1999)
    ∗ vanBuuren and Oudshoorn (1999)
    − Schafer (1997)
    − Rubin (1987)

18. Lab III (Friday, July 16)

    (a) Dependent data

    (b) Missing Data

References

Abbott, Andrew. 1988. "Transcending General Linear Reality." *Sociological Theory* 6(2):169–186.

Abbott, Andrew. 1998. "The Causal Devolution." *Sociological Methods and Research* 27:148–181.

Achen, Christopher H. 1990. "What Does "Explained Variance" Explain?: Reply." *Political Analysis* 2(1):173–184.

Allison, Paul D. 2001. *Missing Data*. Thousand Oaks, CA: Sage.

Andersen, Robert. 2008. *Modern Methods for Robust Regression*. Thousand Oaks, CA: Sage.

Barnes, Grace M, Joseph H. Hoffman, John W Welte, Michael P Farrell and Barbara A Dintcheff. 2006. "Effects of Parental Monitoring and Peer Deviance on Substance Use and Delinquency." *Jounral of Marriage and Family* 68:1084–1104.

Beck, Nathaniel and Jonathan Katz. 1995. "What to do (and not to do) with Time-Series Cross-Section Data." *American Political Science Review* 89(3):634–647.

Beck, Neal and Jonathan Katz. 2007. "Random Coefficient Models for Time-SeriesCross-Section Data: Monte Carlo Experiments." *Political Analysis* 15:182–195.

Berk, Richard A. 1991. "Toward a Methodology for Mere Mortals." *Sociological Methodology* 21:315–324.

Blalock, Hubert M. 1991. "Are There Really Any Constructive Alternatives to Causal Modeling?" *Sociological Methodology* 21:325–335.

Bliese, Paul D. and Robert E. Ployhart. 2002. "Growth Modeling Using Random Coefficient Models: Model Building, Testing and Illustrations." *Organizational Research Methods* 5(4):362–387.

Box, George E. P. 1976. "Science and Statistics." *Journal of the American Statistical Association* 71(356):791–799.

Box, George and P.W. Tidwell. 1962. "Transformation of the Independent Variables." *Technometrics* 4:531–550.

Brambor, Thomas, William Clark and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14(1):63–82.

Braumoeller, Bear F. 2004. "Hypothesis Testing and Multiplicative Interaction Terms." *International Organization* 58(4):807–820.

Burnham, Kenneth P. and David R. Anderson. 2004. "Multimodel Inference: Understanding AIC and BIC in Model Selection." *Sociological Methods and Research* 33(2):261–304.

Cantoni, Gustavo E. and Elvezio Ronchetti. 2001. "Robust Inference for Generalized Linear Models." *Journal of the American Statistical Association* 96:1022–1030.

Clarke, Kevin. 2005. "The Phantom Menace: Omitted Variable Bias in Econometric Research." *Conflict Management and Peace Science* 22(4):341–352.

Cook, R. Dennis and Sanford Weisberg. 1999. *Applied Regression Including Computing and Graphics*. New York: Wiley & Sons, Inc.

Curran, Patrick J. 2003. "Have Multilevel Models Been Structural Equation Models All Along?" *Multivariate Behavioral Research* 38(4):529–569.

Davison, Anthony C. and D.V. Hinkley. 1997. *Bootstrap Methods and Their Applications*. Cambridge: Cambridge University Press.

Downey, Douglas B., Paul T. vonHipple and Broh A. Beckett. 2004. "Are Schools the Great Equalizer? Cognitive Inequality During the Summer Months and the School Year." *American Sociological Review* 69(5):613–635.

Firth, David. 2003. "Overcoming the Reference Category Problem in the Presentation of Statistical Models." *Sociological Methodology* 33:1–18.

Firth, David and Renee X. De Menzes. 2004. "Quasi-Variances." *Biometrika* 91(1):65–80.

Fox, John. 2000*a*. *Multiple and Generalized Nonparametric Regression*. Thousand Oaks: Sage.

Fox, John. 2000*b*. *Nonparametric Simple Regression*. Thousand Oaks: Sage.

Fox, John. 2002. *An R and S-Plus Companion to Applied Regression*. Thousand Oaks: Sage Publications.

Fox, John. 2008. *Applied Regression Analysis and Generalized Linear Models,* 2^{nd} *edition*. Thousand Oaks, CA: Sage, Inc.

Freedman, David A. 1991*a*. "A Rejoinder to Berk, Blalock and Mason." *Sociological Methodology* 21:353–358.

Freedman, David A. 1991*b*. "Statistical Models and Shoe Leather." *Sociological Methodology* 21:291–313.

Gelman, Andrew and Hal Stern. 2006. "The Difference Between 'Significant' and 'Not Significant' is not Itself Statistically Significant." *The American Statistician* 60(4):328–331.

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gill, Jeff. 1999. "The Insignificance of Null Hypothesis Significance Testing." *Political Research Quarterly* 52(3):647–674.

Harvey, Andrew C. 1976. "Estimating Regression Models with Multiplicative Heteroskedasticity." *Econometrica* 44(3):461–465.

Hastie, Trevor J. and Robert J. Tibshirani. 1990. *Generalized Additive Models*. Boca Raton, FL: Chapman & Hall/CRC.

Hsiao, Cheng. 1986. *Analysis of Panel Data*. Cambridge.

Jacoby, William G. 1997. *Statistical Graphics for Univariate and Bivariate Data*. Thousand Oaks, CA: Sage.

Jacoby, William G. 1998. *Statistical Graphics for Visualizing Multivariate Data*. Thousand Oaks, CA: Sage.

Jacoby, William G. 1999. "Levels of Measurement and Political Research: An Optimistic View." *American Journal of Political Science* 43(1):271–301.

Jacoby, William G. 2006. "The Dot Plot: A Graphical Display for Labeled QUantitative Values." *The Political Methodologist* 14(1):6–14.

Jasso, Guillermina. 1985. "Marital Coital Frequency and the Passage of Time: Estimating the Separate Effects of Spouses' Ages and Marital Duration, Birth and Marriage Cohorts, and Period Influences." *American Sociological Review* 50(2):224–241.

Jasso, Guillermina. 1996. "Is It Outlier Deletion or is it Sample Truncation? Notes on Science and Sexuality." *American Sociological Review* 51(5):738–742.

Kahn, Joan R. and J. Richard Udry. 1986. "Marital Coital Frequency: Unnoticed Outliers and Unspecified Interactions Lead to Erroneous Conclusions." *American Sociological Review* 51(5):734–737.

Kam, Cindy and Robert J. Franzese. 2007. *Modeling and Interpreting Interactive Hypotheses in Regression Analyses*. Ann Arbor: University of Michigan Press.

Kastellec, Jonathan P and Eduardo L. Leoni. 2007. "Using Graphs Instead of Tables in Political Science." *Perspectives on Politics* 5(4):755–771.

Keele, Luke J. 2008. *Semiparametric Regression for the Social Sciences*. New York: Wiley & Sons, Inc.

Kim, ChangHwan and Arthur Sakamoto. 2008. "The Rise of Intra-Operational Wage Inequality in the United States, 1983-2002." *American Sociological Review* 73(1):129–157.

King, Gary. 1990. "Stochastic Variation: A Comment on Lewis-Beck and Skalaban's 'The R-Squared'." *Political Analysis* 2(1):185–200.

Leamer, Edward E. 1974. "False Models and Post-Data Model Construction." *Journal of the American Statistical Association* 69(345):122–131.

Leamer, Edward E. 1978. *Specification Searches: Ad Hoc Inference With Nonexperimental Data*. New York: Wiley & Sons, Inc.

Leamer, Edward E. 1983. "Let's Take the Con Out of Econometrics." *The American Economic Review* 73(1):31–43.

Leamer, Edward E. and Herman Leonard. 1983. "Reporting the Fragility of Regression Estimates." *The Review of Economics and Statistics* 65(2):306–317.

Lewis-Beck, Michael S. and Andrew Skalaban. 1990. "The R-squared: Some Straight Talk." *Political Analysis* 2(1):153–171.

Long, J. Scott and Laurie H. Ervin. 2000. "Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model." *The American Statistician* 54(3):217–224.

Mallows, Colin L. 1974. "Some Comments on C_p." *Technometrics* 15:661–675.

Mason, William M. 1991. "Freedman is Right as Far as He Goes, but THere is More, and It's Worse. Statisticians Could Help." *Sociological Methodology* 21:337–357.

Miller, Alan. 2002. *Subset Selection in Regression,* 2^{nd} *edition*. Boca Raton, FL: Chapman & Hall/CRC.

Miller, Alan J. 1984. "Selection of Subsets of Regression Variables." *Journal of the Royal Statistical Society, Series A* 147(3):389–425.

Murrell, Paul. 2006. *R Graphics*. Boca Raton, FL: Chapman & Hall/CRC.

Pinheiro, Jose C. and Douglas M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag.

Plumper, Thomas and Vera E. Troeger. 2007. "Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects." *Political Analysis* 15(2):124–139.

Raudenbush, Stephen and Anthony Bryk. 2002. *Hierarchical Linear Models*. Thousand Oaks, CA: Sage.

Rousseeuw, Peter J and Annick M. Leroy. 1987. *Robust Regression and Outlier Detection*. New York: Wiley & Sons, Inc.

Rubin, Donald. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: J. Wiley and Sons.

Schafer, Joseph L. 1997. *Analysis of Incomplete Multivariate Data*. New York: Chapman and Hall/CRC.

Schafer, Joseph L. 1999. "Multiple Imputation: A Primer." *Statistical Methods in Medical Research* 8:3–15.

Silber, Jeffrey H., Paul R. Rosenbaum and Richard N. Ross. 1995. "Comparing the Contributions of Groups of Predictors: Which Outcomes Vary with Hospital Rather Than Patient Characteristics." *Journal of the American Statistical Association* 90(429):7–18.

Singer, Judith D. and John B Willett. 2003. *Applied Longitudinal Data Analysis*. New York: Oxford University Press.

Snijders, Tom A. B. and Roel Bosker. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage.

vanBuuren, Stef and Karin Oudshoorn. 1999. Flexible Multivariate Imputation by MICE. Technical report TNO Prevention and Health.

Venables, William N. and Brian D. Ripley. 2002. *Modern Applied Statistics with S,* 4^{th} *edition*. 3 ed. New York: Springer.

Wilson, Sven E. and Daniel M. Butler. 2007. "A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications." *Political Analysis* 15(2):101–123.

Wood, Simon. 2006. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC.

Wooldridge, Jeffrey. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Yohai, Victor J. 1987. "High Breakdown Point and High Efficiency Robust Estimates for Regression." *Annals of Statistics* 15:642–656.

ICPSR Summer Program in Quantitative Methods of Social Research.Â Empirical Modeling of Social Science Theory: Advanced Topics. Introduction to the R Statistical Computing Environment. Statistics and Data Analysis I: Introduction. Regression Analysis III: Advanced Methods. Pagination. Next page Next â€º. Regression Analysis III: Advanced Methods Syllabus-Summer 2010. Dave Armstrong. Matthew Painter. The Regression III course takes a considerably different form than the first two re-gression courses at the Summer Program. This course will hopefully prepare you for the things you will encounter when you (attempt to) publish quantitative work with linear models. Initial linear model classes focus on the assumptions and theoretical consider-ations of linear models and generally walk you through We will analyze the strengths and weaknesses of these methods. Applications are drawn from various elds including political science, public policy, economics, and sociology. The class is open to qualied students from other departments and undergraduates. Prerequisites A willingness to work hard on possibly unfamiliar material. In addition to introductory statistics and proba-bility, the course assumes a good knowledge of linear regression meaning that you should have taken at least one graduate class on this subject (such as 17.800).Â Class Requirements Reading The syllabus lists the required readings for every week. This required reading should be completed prior to lecture in a given week. Students are expected to read the material very carefully.