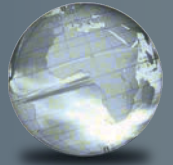


GLOBAL
EDITION



Econometric Analysis

EIGHTH EDITION

William H. Greene



For Margaret and Richard Greene

Director of Global Content Strategy, Business and HSS:
Lacey Vitetta
Director of Product Management, Business and Economics:
Adrienne D' Ambrosio
Senior Sponsoring Editor: Neeraj Bhalla
Editorial Assistant: Courtney Paganelli
Managing Editor, Global Edition: Yajnaseni Das
Project Editor, Global Edition: Paromita Banerjee
Vice President, Product Marketing: Roxanne McCarley
Director of Strategic Marketing: Brad Parkins
Strategic Marketing Manager: Deborah Strickland
Product Marketer: Tricia Murphy
Field Marketing Manager: Ramona Elmer
Product Marketing Assistant: Jessica Quazza
Vice President, Production and Digital Studio, Arts and
Business: Etain O'Dea
Director of Production, Business: Jeff Holcomb
Managing Producer, Business: Alison Kalil

Content Producer: Sugandh Juneja
Content Producer, Global Edition: Isha Sachdeva
Operations Specialist: Carol Melville
Manufacturing Controller, Global Edition: Angela Hawksbee
Creative Director: Blair Brown
Manager, Learning Tools: Brian Surette
Content Developer, Learning Tools: Lindsey Sloan
Managing Producer, Digital Studio, Arts and Business: Diane Lombardo
Digital Studio Producer: Melissa Honig
Digital Studio Producer: Alana Coles
Digital Content Team Lead: Noel Lotz
Digital Content Project Lead: Courtney Kamauf
Full-Service Project Management and Composition:
SPi Global
Interior Design: SPi Global
Cover Design: Lumina Datamatics
Cover Art: Login/Shutterstock

Microsoft and/or its respective suppliers make no representations about the suitability of the information contained in the documents and related graphics published as part of the services for any purpose. All such documents and related graphics are provided "as is" without warranty of any kind. Microsoft and/or its respective suppliers hereby disclaim all warranties and conditions with regard to this information, including all warranties and conditions of merchantability, whether express, implied or statutory, fitness for a particular purpose, title and non-infringement. In no event shall Microsoft and/or its respective suppliers be liable for any special, indirect or consequential damages or any damages whatsoever resulting from loss of use, data or profits, whether in an action of contract, negligence or other tortious action, arising out of or in connection with the use or performance of information available from the services.

The documents and related graphics contained herein could include technical inaccuracies or typographical errors. Changes are periodically added to the information herein. Microsoft and/or its respective suppliers may make improvements and/or changes in the product(s) and/or the program(s) described herein at any time. Partial screen shots may be viewed in full within the software version specified.

Acknowledgments of third-party content appear on the appropriate page within the text.

Pearson Education Limited

KAO TWO
KAO Park
Hockham Way
Harlow
Essex
CM17 9SR

and Associated Companies throughout the world

Visit us on the World Wide Web at: www.pearsonglobaleditions.com

© Pearson Education Limited, 2020

The rights of William H. Greene to be identified as the author of this work have been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Authorized adaptation from the United States edition, entitled *Econometric Analysis*, 8th Edition, ISBN 978-0-13-446136-6 by William H. Greene, published by Pearson Education © 2018.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without either the prior written permission of the publisher or a license permitting restricted copying in the United Kingdom issued by the Copyright Licensing Agency Ltd, Saffron House, 6–10 Kirby Street, London EC1N 8TS. This publication is protected by copyright, and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise. For information regarding permissions, request forms, and the appropriate contacts within the Pearson Education Global Rights and Permissions department, please visit www.pearsoned.com/permissions/.

This eBook is a standalone product and may or may not include all assets that were part of the print version. It also does not provide access to other Pearson digital products like MyLab and Mastering. The publisher reserves the right to remove any material in this eBook at any time.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN 10: 1-292-23113-0

ISBN 13: 978-1-292-23113-6

eBook ISBN 13: 978-1-292-23115-0

Typeset in Times Ten LT Std by SPi Global

With this formulation, all of the parameters are identified. This is an example of an exactly identified model. An additional variation is worth a look. Suppose that a second variable, w (weather), appears in the supply equation,

$$q = \alpha_0 + \alpha_1 p + \alpha_2 x + \varepsilon_1,$$

$$q = \beta_0 + \beta_1 p + \beta_2 z + \beta_3 w + \varepsilon_2.$$

You can easily verify that, the reduced form matrix is the same as the previous one, save for an additional row that contains $[\alpha_1 \beta_3 / \Lambda, \beta_3 / \Lambda]$. This implies that there is now a second solution for $\alpha_1, \pi_{41} / \pi_{42}$. The two solutions, this and π_{31} / π_{32} , will be different. This model is overidentified. There is more information in the sample and theory than is needed to deduce the structural parameters.

Some equation systems are identified and others are not. The formal mathematical conditions under which an equation in a system is identified turns on two results known as the rank and order conditions. The *order condition* is a simple counting rule. It requires that the number of exogenous variables that appear elsewhere in the equation system must be at least as large as the number of endogenous variables in the equation. (Other specific restrictions on the parameters will be included in this count—note that an “exclusion restriction” is a type of linear restriction.) We used this rule when we constructed the IV estimator in Chapter 8. In that setting, we required our model to be at least *identified* by requiring that the number of instrumental variables not contained in \mathbf{X} be at least as large as the number of endogenous variables. The correspondence of that single equation application with the condition defined here is that the rest of the equation system is the source of the instrumental variables. One simple order condition for identification of an equation system is that each equation contain “its own” exogenous variable that does not appear elsewhere in the system.

The **order condition** is necessary for identification; the **rank condition** is sufficient. The equation system in (10-37) in structural form is $\mathbf{y}'\Gamma = -\mathbf{x}'\mathbf{B} + \boldsymbol{\varepsilon}'$. The reduced form is $\mathbf{y}' = \mathbf{x}'(-\mathbf{B}\Gamma^{-1}) + \boldsymbol{\varepsilon}'\Gamma^{-1} = \mathbf{x}'\Pi + \mathbf{v}'$. The way we are going to deduce the parameters in $(\Gamma, \mathbf{B}, \Sigma)$ is from the reduced form parameters (Π, Ω) . For the j th equation, the solution is contained in $\Pi\Gamma_j = -\mathbf{B}_j$, where Γ_j contains all the coefficients in the j th equation that multiply endogenous variables. One of these coefficients will equal one, usually some will equal zero, and the remainder are the nonzero coefficients on endogenous variables in the equation, \mathbf{Y}_j [these are denoted γ_j in (10-41) following]. Likewise, \mathbf{B}_j contains the coefficients in equation j on all exogenous variables in the model—some of these will be zero and the remainder will multiply variables in \mathbf{X}_j , the exogenous variables that appear in this equation [these are denoted β_j in (10-41) following]. The empirical counterpart will be $\mathbf{Pc}_j = \mathbf{b}_j$, where \mathbf{P} is the estimated reduced form, $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, and \mathbf{c}_j and \mathbf{b}_j will be the estimates of the j th columns of Γ and \mathbf{B} . The rank condition ensures that there is a solution to this set of equations. In practical terms, the rank condition is difficult to establish in large equation systems. Practitioners typically take it as a given. In small systems, such as the two-equation systems that dominate contemporary research, it is trivial, as we examine in the next example. We have already used the rank condition in Chapter 8, where it played a role in the relevance condition for instrumental variable estimation. In particular, note after the statement of the assumptions for instrumental variable estimation, we assumed $\text{plim}(1/T)\mathbf{Z}'\mathbf{X}$ is a matrix with rank K . (This condition is often labeled the *rank condition* in contemporary applications. It not identical, but it is sufficient for the condition mentioned here.)

Example 10.7 The Rank Condition and a Two-Equation Model

The following two-equation recursive model provides what is arguably the platform for much of contemporary econometric analysis. The main equation of interest is

$$y = \gamma f + \beta x + \varepsilon.$$

The variable f is endogenous (it is correlated with ε); x is exogenous (it is uncorrelated with ε). The analyst has in hand an instrument for f , z . The instrument, z , is relevant, in that in the auxiliary equation,

$$f = \lambda x + \delta z + w,$$

δ is not zero. The exogeneity assumption is $E[\varepsilon z] = E[wz] = 0$. Note that the source of the endogeneity of f is the assumed correlation of w and ε . For purposes of the exercise, assume that $E[xz] = 0$ and the data satisfy $\mathbf{x}'\mathbf{z} = 0$ —this actually loses no generality. In this two-equation model, the second equation is already in reduced form; x and z are both exogenous. It follows that λ and δ are estimable by least squares. The estimating equations for (γ, β) are

$$\mathbf{P}\gamma_1 = \begin{bmatrix} \mathbf{x}'\mathbf{x} & \mathbf{x}'\mathbf{z} \\ \mathbf{z}'\mathbf{x} & \mathbf{z}'\mathbf{z} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x}'\mathbf{y} & \mathbf{x}'\mathbf{f} \\ \mathbf{z}'\mathbf{y} & \mathbf{z}'\mathbf{f} \end{bmatrix} \begin{pmatrix} 1 \\ -\gamma \end{pmatrix} = \begin{bmatrix} \mathbf{x}'\mathbf{y}/\mathbf{x}'\mathbf{x} & \mathbf{x}'\mathbf{f}/\mathbf{x}'\mathbf{x} \\ \mathbf{z}'\mathbf{y}/\mathbf{z}'\mathbf{z} & \mathbf{z}'\mathbf{f}/\mathbf{z}'\mathbf{z} \end{bmatrix} \begin{pmatrix} 1 \\ -\gamma \end{pmatrix} = \beta_j = \begin{pmatrix} \beta \\ 0 \end{pmatrix}.$$

The solutions are $\gamma = (\mathbf{z}'\mathbf{y}/\mathbf{z}'\mathbf{f})$ and $\beta = (\mathbf{x}'\mathbf{y}/\mathbf{x}'\mathbf{x} - (\mathbf{z}'\mathbf{y}/\mathbf{z}'\mathbf{f})\mathbf{x}'\mathbf{f}/\mathbf{x}'\mathbf{x})$. Because $\mathbf{x}'\mathbf{x}$ cannot equal zero, the solution depends on $(\mathbf{z}'\mathbf{f}/\mathbf{z}'\mathbf{z})$ not equal to zero—formally that this part of the reduced form coefficient matrix have rank $M = 1$, which would be the rank condition. Note that the solution for γ is the instrumental variable estimator, with z as instrument for f . (The simplicity of this solution turns on the assumption that $\mathbf{x}'\mathbf{z} = 0$. The algebra gets a bit more complicated without it, but the conclusion is the same.)

The rank condition is based on the exclusion restrictions in the model—whether the exclusion restrictions provide enough information to identify the coefficients in the j th equation. Formally, the idea can be developed thusly. With the j th equation written as in (10-41), we call \mathbf{X}_j the *included exogenous variables*. The remaining excluded exogenous variables are denoted \mathbf{X}_j^* . The M_j variables \mathbf{Y}_j in (10-41) are the included endogenous variables. With this distinction, we can write the M_j reduced forms for \mathbf{Y}_j as $\bar{\Pi}_j = \begin{bmatrix} \Pi_j \\ \Pi_j^* \end{bmatrix}$. The rank condition (which we state without proof) is that the rank of the lower part of the $M_j \times (K_j + K_j^*)$ matrix, $\bar{\Pi}_j$, equal M_j . In the preceding example, in the first equation, \mathbf{Y}_j is f , $M_j = 1$, \mathbf{X}_j is x , \mathbf{X}_j^* is z , and $\bar{\Pi}_j$ is estimated by the regression of f on x and z ; Π_j is the coefficient on x and Π_j^* is the coefficient on z . The rank condition we noted earlier is that what is estimated by $\mathbf{z}'\mathbf{f}/\mathbf{z}'\mathbf{z}$, which would correspond to Π_j^* not equal zero, meaning that it has rank 1.

Casual statements of the rank condition based on an IV regression of a variable \mathbf{y}_{IV} on $(M_j + K_j)$ endogenous and exogeneous variables in \mathbf{X}_{IV} , using $K_j + K_j^*$ exogenous and instrumental variables in \mathbf{Z}_{IV} (in the most familiar cases, $M_j = K_j^* = 1$), state that the rank requirement is that $(\mathbf{Z}_{IV}'\mathbf{X}_{IV}/T)$ be nonsingular. In the notation we are using here, \mathbf{Z}_{IV} would be $\mathbf{X} = (\mathbf{X}_j, \mathbf{X}_j^*)$ and \mathbf{X}_{IV} would be $(\mathbf{X}_j, \mathbf{Y}_j)$. This nonsingularity would correspond to full rank of $\text{plim}[(\mathbf{X}'\mathbf{X}/T)]$ times $\text{plim}[(\mathbf{X}'\mathbf{X}^*/T, \mathbf{X}'\mathbf{Y}_j/T)]$ because $\text{plim}[(\mathbf{X}'\mathbf{X}/T)] = \mathbf{Q}$, which is nonsingular [see (10-40)]. The first K_j columns of this matrix are the last K_j columns of an identity matrix, which have rank K_j . The last M_j columns are estimates of $\mathbf{Q}\bar{\Pi}_j$, which we require to have rank M_j , so the requirement is that $\bar{\Pi}_j$ have rank M_j . But, if $K_j^* \geq M_j$ (the order condition), then all that is needed is $\text{rank}(\Pi_j^*) = M_j$, so, in practical terms, the casual statement is correct. It is stronger than necessary; the formal mathematical condition is only that the lower half of the matrix must have rank M_j , but the practical result is much easier to visualize.

It is also easy to verify that the rank condition requires that the predictions of \mathbf{Y}_j using $(\mathbf{X}_j, \mathbf{X}_j^*)\bar{\Pi}_j$ be linearly independent. Continuing this line of thinking, if we use 2SLS, the rank condition requires that the predicted values of the included endogenous variables not be collinear, which makes sense.

10.4.4 SINGLE EQUATION ESTIMATION AND INFERENCE

For purposes of estimation and inference, we write the model in the way that the researcher would typically formulate it,

$$\begin{aligned} \mathbf{y}_j &= \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{Y}_j\boldsymbol{\gamma}_j + \boldsymbol{\varepsilon}_j \\ &= \mathbf{Z}_j\boldsymbol{\delta}_j + \boldsymbol{\varepsilon}_j, \end{aligned} \quad (10-41)$$

where \mathbf{y}_j is the “dependent variable” in the equation, \mathbf{X}_j is the set of exogenous variables that appear in the j th equation—note that this is not all the variables in the model—and $\mathbf{Z}_j = (\mathbf{X}_j, \mathbf{Y}_j)$. The full set of exogenous variables in the model, including \mathbf{X}_j and variables that appear elsewhere in the model (including a constant term if any equation includes one), is denoted \mathbf{X} . For example, in the supply/demand model in Example 10.6, the full set of exogenous variables is $\mathbf{X} = (\mathbf{1}, \mathbf{x}, \mathbf{z})$, while $\mathbf{X}_{Demand} = (\mathbf{1}, \mathbf{x})$ and $\mathbf{X}_{Supply} = (\mathbf{1}, \mathbf{z})$. Finally, \mathbf{Y}_j is the endogenous variables that appear on the right-hand side of the j th equation. Once again, this is likely to be a subset of the endogenous variables in the full model. In Example 10.6, $\mathbf{Y}_j = (\text{price})$ in both cases.

There are two approaches to estimation and inference for simultaneous equations models. **Limited information estimators** are constructed for each equation individually. The approach is analogous to estimation of the seemingly unrelated regressions model in Section 10.2 by least squares, one equation at a time. **Full information estimators** are used to estimate all equations simultaneously. The counterpart for the seemingly unrelated regressions model is the feasible generalized least squares estimator discussed in Section 10.2.3. The major difference to be accommodated at this point is the endogeneity of \mathbf{Y}_j in (10-41).

The equation in (10-41) is precisely the model developed in Chapter 8. Least squares will generally be unsuitable as it is inconsistent due to the correlation between \mathbf{Y}_j and $\boldsymbol{\varepsilon}_j$. The usual approach will be two-stage least squares as developed in Sections 8.3.2 through 8.3.4. The only difference between the case considered here and that in Chapter 8 is the source of the instrumental variables. In our general model in Chapter 8, the source of the instruments remained somewhat ambiguous; the overall rule was “outside the model.” In this setting, the instruments come from elsewhere in the model—that is, “not in the j th equation.” For estimating the linear simultaneous equations model, the most common estimator is

$$\begin{aligned} \hat{\boldsymbol{\delta}}_{j, 2 \text{ SLS}} &= [\hat{\mathbf{Z}}_j' \hat{\mathbf{Z}}_j]^{-1} \hat{\mathbf{Z}}_j' \mathbf{y}_j \\ &= [(\mathbf{Z}_j' \mathbf{X})(\mathbf{X}' \mathbf{X})^{-1}(\mathbf{X}' \mathbf{Z}_j)]^{-1} (\mathbf{Z}_j' \mathbf{X})(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_j, \end{aligned} \quad (10-42)$$

where all columns of $\hat{\mathbf{Z}}_j'$ are obtained as predictions in a regression of the corresponding column of \mathbf{Z}_j on \mathbf{X} . This equation also results in a useful simplification of the estimated asymptotic covariance matrix,

$$\text{Est.Asy.Var}[\hat{\boldsymbol{\delta}}_{j, 2 \text{ SLS}}] = \hat{\sigma}_{jj}(\hat{\mathbf{Z}}_j' \hat{\mathbf{Z}}_j)^{-1}.$$

It is important to note that σ_{jj} is estimated by

$$\hat{\sigma}_{jj} = \frac{(\mathbf{y}_j - \mathbf{Z}_j \hat{\boldsymbol{\delta}}_j)'(\mathbf{y}_j - \mathbf{Z}_j \hat{\boldsymbol{\delta}}_j)}{T}, \quad (10-43)$$

using the original data, not $\hat{\mathbf{Z}}_j$.

Note the role of the order condition for identification in the two-stage least squares estimator. Formally, the order condition requires that the number of exogenous variables that appear elsewhere in the model (not in this equation) be at least as large as the number of endogenous variables that appear in this equation. The implication will be that we are going to predict $\mathbf{Z}_j = (\mathbf{X}_j, \mathbf{Y}_j)$ using $\mathbf{X} = (\mathbf{X}_j, \mathbf{X}_j^*)$. In order for these predictions to be linearly independent, there must be at least as many variables used to compute the predictions as there are variables being predicted. Comparing $(\mathbf{X}_j, \mathbf{Y}_j)$ to $(\mathbf{X}_j, \mathbf{X}_j^*)$, we see that there must be at least as many variables in \mathbf{X}_j^* as there are in \mathbf{Y}_j , which is the order condition. The practical rule of thumb that every equation have at least one variable in it that does not appear in any other equation will guarantee this outcome.

Two-stage least squares is used nearly universally in estimation of linear simultaneous equation models—for precisely the reasons outlined in Chapter 8. However, some applications (and some theoretical treatments) have suggested that the **limited information maximum likelihood (LIML) estimator** based on the normal distribution may have better properties. The technique has also found recent use in the analysis of weak instruments. A result that emerges from the derivation is that the LIML estimator has the same asymptotic distribution as the 2SLS estimator, and the latter does not rely on an assumption of normality. This raises the question why one would use the LIML technique given the availability of the more robust (and computationally simpler) alternative. Small sample results are sparse, but they would favor 2SLS as well.²⁵ One significant virtue of LIML is its invariance to the normalization of the equation. Consider an example in a system of equations,

$$y_1 = y_2\gamma_2 + y_3\gamma_3 + x_1\beta_1 + x_2\beta_2 + \varepsilon_1.$$

An equivalent equation would be

$$\begin{aligned} y_2 &= y_1(1/\gamma_2) + y_3(-\gamma_3/\gamma_2) + x_1(-\beta_1/\gamma_2) + x_2(-\beta_2/\gamma_2) + \varepsilon_1(-1/\gamma_2) \\ &= y_1\tilde{\gamma}_1 + y_3\tilde{\gamma}_3 + x_1\tilde{\beta}_1 + x_2\tilde{\beta}_2 + \tilde{\varepsilon}_1. \end{aligned}$$

The parameters of the second equation can be manipulated to produce those of the first. But, as you can easily verify, the 2SLS estimator is not invariant to the normalization of the equation—2SLS would produce numerically different answers. LIML would give the same numerical solutions to both estimation problems suggested earlier. A second virtue is LIML's better performance in the presence of weak instruments.

The LIML, or **least variance ratio** estimator, can be computed as follows.²⁶ Let

$$\mathbf{W}_j^0 = \mathbf{E}_j^{0'} \mathbf{E}_j^0, \quad (10-44)$$

where

$$\mathbf{Y}_j^0 = [\mathbf{y}_j, \mathbf{Y}_j],$$

and

$$\mathbf{E}_j^0 = \mathbf{M}_j \mathbf{Y}_j^0 = [\mathbf{I} - \mathbf{X}_j(\mathbf{X}_j' \mathbf{X}_j)^{-1} \mathbf{X}_j'] \mathbf{Y}_j^0. \quad (10-45)$$

²⁵See Phillips (1983).

²⁶The LIML estimator was derived by Anderson and Rubin (1949, 1950). [See, also, Johnston (1984).] The much simpler and equally efficient two-stage least squares estimator remains the estimator of choice.

Each column of \mathbf{E}_j^0 is a set of least squares residuals in the regression of the corresponding column of \mathbf{Y}_j^0 on \mathbf{X}_j , that is, only the exogenous variables that appear in the j th equation. Thus, \mathbf{W}_j^0 is the matrix of sums of squares and cross products of these residuals. Define

$$\mathbf{W}_j^1 = \mathbf{E}_j^1 \mathbf{E}_j^{1'} = \mathbf{Y}_j^{0'} [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \mathbf{Y}_j^0. \quad (10-46)$$

That is, \mathbf{W}_j^1 is defined like \mathbf{W}_j^0 except that the regressions are on all the x 's in the model, not just the ones in the j th equation. Let

$$\lambda_1 = \text{smallest characteristic root of } (\mathbf{W}_j^1)^{-1}\mathbf{W}_j^0. \quad (10-47)$$

This matrix is asymmetric, but all its roots are real and greater than or equal to 1. [Depending on the available software, it may be more convenient to obtain the identical smallest root of the symmetric matrix $\mathbf{D} = (\mathbf{W}_j^1)^{-1/2}\mathbf{W}_j^0(\mathbf{W}_j^1)^{-1/2}$.] Now partition \mathbf{W}_j^0 into $\mathbf{W}_j^0 = \begin{bmatrix} \mathbf{w}_{jj}^0 & \mathbf{w}_j^{0'} \\ \mathbf{w}_j^0 & \mathbf{W}_{jj}^0 \end{bmatrix}$ corresponding to $[\mathbf{y}_j, \mathbf{Y}_j]$, and partition \mathbf{W}_j^1 likewise. Then, with these parts in hand,

$$\hat{\gamma}_{j, \text{LIML}} = [\mathbf{W}_{jj}^0 - \lambda_1 \mathbf{W}_{jj}^1]^{-1}(\mathbf{w}_j^0 - \lambda_1 \mathbf{w}_j^1) \quad (10-48)$$

and

$$\hat{\beta}_{j, \text{LIML}} = (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'(\mathbf{y}_j - \mathbf{Y}_j\hat{\gamma}_{j, \text{LIML}}).$$

Note that β_j is estimated by a simple least squares regression. [See (3-18).] The asymptotic covariance matrix for the LIML estimator is identical to that for the 2SLS estimator.

Example 10.8 Simultaneity in Health Production

Example 7.1 analyzed the incomes of a subsample of Riphahn, Wambach, and Million's (2003) data on health outcomes in the German Socioeconomic Panel. Here we continue Example 10.4 and consider a Grossman (1972) style model for health and incomes. Our two-equation model is

$$\begin{aligned} \text{Health Satisfaction} = & \alpha_1 + \gamma_1 \text{In Income} + \alpha_2 \text{Female} + \alpha_3 \text{Working} + \alpha_4 \text{Public} + \alpha_5 \text{Add On} \\ & + \alpha_6 \text{Age} + \varepsilon_H, \end{aligned}$$

$$\begin{aligned} \text{In Income} = & \beta_1 + \gamma_2 \text{Health Satisfaction} + \beta_2 \text{Female} + \beta_3 \text{Education} + \beta_4 \text{Married} \\ & + \beta_5 \text{HHKids} + \beta_6 \text{Age} + \varepsilon_I. \end{aligned}$$

For purposes of this application, we avoid panel data considerations by examining only the 1994 wave (cross section) of the data, which contains 3,377 observations. The health outcome variable is *Self Assessed Health Satisfaction* (HSAT). Whether this variable actually corresponds to a commonly defined objective measure of health outcomes is debateable. We will treat it as such. Second, the variable is a scale variable, coded in this data set 0 to 10. [In more recent versions of the GSOEP data, and in the British (BHPS) and Australian (HILDA) counterparts, it is coded 0 to 4.] We would ordinarily treat such a variable as a discrete ordered outcome, as we do in Examples 18.14 and 18.15. We will treat it as if it were continuous in this example, and recognize that there is likely to be some distortion in the measured effects that we are interested in. *Female*, *Working*, *Married*, and *HHkids* are dummy variables, the last indicating whether there are children living in the household. *Education* and *Age* are in years. *Public* and *AddOn* are dummy variables that indicate whether the individual takes up the public health insurance and, if so, whether he or she also takes up the additional