Chapter: Simultaneous Equations Models and Generalized Method of Moment Estimation

5.1 Simultaneous Equations Models

5.1.1 Model structures

Model: (Structural form; structural equations)

\[ y_t' \Gamma + x_t' B = \epsilon_t', \]

where

\[ \Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1m} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \cdots & \gamma_{mm} \end{pmatrix}, \quad B = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{G1} & \beta_{G2} & \cdots & \beta_{Gm} \end{pmatrix}, \]

and \( x_t' = (x_{1t}, \ldots, x_{Gt}) \) and \( y_t' = (y_{1t}, \ldots, y_{mt}) \).

- Reduced form equations

\[ y_t' = x_t' \Pi + v_t', \]

where \( \Pi = -B \Gamma^{-1} \) and \( v_t' = \epsilon_t' \Gamma^{-1} \).

- Variable classification: \( y_t \) jointly dependent or endogenous variables – determined inside the model; \( x_t \) exogenous variables – determined outside the model. For the exogenous variables, it is usually assumed that \( \text{plim} B = 0 \).

- Reduced form estimation: \( \Pi \) and the variance \( \Omega \) of \( v_t \) can be consistently estimated by OLS regression of \( y \) on \( x \).

5.1.2 Identification of structural parameters

- Identification problem: Can the structural parameters be deduced from the reduced form parameters?

\[ \Pi = -B \Gamma^{-1}, \quad \Omega = (\Gamma^{-1})' \Sigma (\Gamma^{-1}) \]

where \( \Sigma \) is the variance of \( \epsilon_t \).

1) If there were no nonsample restrictions on \( B \) and \( \Gamma \), they are not identifiable. For any nonsingular matrix \( F \), the system \( y_t' \Gamma + x_t' \tilde{B} = \tilde{\epsilon} \) where \( \tilde{\Gamma} = \Gamma F, \tilde{B} = BF, \) and \( \tilde{\epsilon} = \epsilon' F \). The reduced forms for the two systems are the same (with the same reduced form parameters).

2) Exclusion restrictions:

\[ y_j = Y_j' \gamma_j + Y_j'^* \gamma_j^* + x_j' \beta_j + x_j'^* \beta_j^* + \epsilon_j, \]

where \( \gamma_j^* = 0 \) and \( \beta_j^* = 0 \). Thus, in \( y_j' \Gamma_j + x_j' B_j = \epsilon_j, \)

\[ \Gamma_j = [1, -\gamma_j', 0]', \quad B_j = [-\beta_j', 0']. \]

Write the reduced form \( y' = x' \Pi + v \) as

\[ [y_j' \ Y_j'^*] = [x_j' \ x_j'^*] \left( \begin{pmatrix} \pi_j \\ \pi_j^* \end{pmatrix} \begin{pmatrix} \Pi_j \\ \Pi_j^* \end{pmatrix} \right) + [v_j' \ V_j'^*]. \]

Since \( \Pi \Gamma_j + B_j = 0 \) (implied from \( \Pi \Gamma = -B \)),

\[ \left( \begin{pmatrix} \pi_j \\ \pi_j^* \end{pmatrix} \begin{pmatrix} \Pi_j \\ \Pi_j^* \end{pmatrix} \right) \left( \begin{pmatrix} 1 \\ -\gamma_j \end{pmatrix} \right) = \left( \begin{pmatrix} \beta_j \\ 0 \end{pmatrix} \right). \]

This implies that

\[ \pi_j - \Pi_j \gamma_j = \beta_j, \quad \pi_j^* - \Pi_j^* \gamma_j = 0. \]
So the problem is reduced to solve $\gamma_j$ uniquely from

$$\Pi_j^* \gamma_j = \pi_j^*.$$ 

Let the number of excluded exogenous variables in the $j$th equation be $G_j^*$, and the number of included endogenous variables on the right hand side (i.e., dimension of $Y_j$) be $m_j$. Then $\Pi_j^*$ is a $G_j^* \times m_j$ matrix.

- **Order condition for identification** (of the $j$th eq.): $G_j^* \geq m_j$. The number of exogenous variables excluded from equation $j$ must be at least as large as the number of endogenous variables included in the $j$th equation. This is a sufficient condition (but not necessary).

- **Rank condition for identification** (of the $j$th eq.):

  $$\text{rank}[\Pi_j^*] = m_j.$$

  This is a necessary and sufficient condition for identification.

  - **Equivalent rank condition**: Partition the structural equation matrix into

    $$\begin{pmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \end{pmatrix} = \begin{pmatrix} 1 \\ -\gamma_j \\ 0 \\ -\beta_j \\ 0 \end{pmatrix}.$$

  An equivalent rank condition is $\text{rank} \left( \begin{pmatrix} A_3 \\ A_5 \end{pmatrix} \right) = m_j - 1$.

- **Terminology on identification**:
  1. \textit{Underidentified}. $G_j^* < m_j$ or rank condition fails.
  2. \textit{Exactly identified}. $G_j^* = m_j$ and rank condition is met.
  3. \textit{Overidentified}. $G_j^* > m_j$ and rank condition is met.

**A recursive model with uncorrelated disturbances (a fully recursive system) is identifiable.**

### 5.1.3 Estimation

Consider the estimation of the (identifiable) $j$th equation:

$$y_j = Y_j \gamma_j + X_j \beta_j + \epsilon_j = Z_j \delta_j + \epsilon_j,$$

where $Z_j = (Y_j, X_j)$ and $\delta_j = (\gamma_j', \beta_j')'$.

- **OLSE of $\delta_j$** is inconsistent because $\epsilon_j$ is correlated with $Y_j$ (and hence $Z_j$).

- **Instrumental variable estimators (IV)**:

  Let $W_j$ be a matrix satisfying the properties that $\text{plim} \left( \frac{W_j^T Z_j}{T} \right) = \Sigma_w$ - a finite nonsingular matrix; $\text{plim} \left( \frac{W_j^T}{T} \right) = 0$; and $\text{plim} \left( \frac{W_j^T W_j}{T} \right) = \Sigma_{ww}$ - a positive definite matrix.

  An IV estimator with IV matrix $W_j$ is

  $$\hat{\delta}_{j,IV} = \left( W_j^T Z_j \right)^{-1} W_j^T y_j.$$

  1.) The $\hat{\delta}_{j,IV}$ is consistent and its asymptotic variance matrix is

  $$\text{Asy.Var}[\hat{\delta}_{j,IV}] = \sigma_{jj}^2 T \text{plim} \left[ \frac{W_j^T Z_j}{T} \right]^{-1} \left[ \frac{W_j^T W_j}{T} \right] \left[ \frac{Z_j^T W_j}{T} \right]^{-1}.$$

  2.) $\hat{\sigma}_{jj} = (Y_j - Z_j \hat{\delta}_{j,IV})(Y_j - Z_j \hat{\delta}_{j,IV})/T$ is a consistent estimate of $\sigma_{jj}$.

- **Two-stage Least Squares estimator**:

  $$\hat{\delta}_{j,2SLS} = \left( \hat{Y}_j Y_j \hat{X}_j X_j \hat{X}_j X_j \right)^{-1} \left( \hat{Y}_j y_j \hat{X}_j y_j \right),$$

2
where $\hat{Y}_j = X[(X'X)^{-1}X'Y_j]$. It corresponds to an IV with $W_j = [\hat{Y}_j, X_j]$. 

1.) An equivalent expression is 

$$
\delta_{j,2SLS} = (\hat{Z}_j'\hat{Z}_j)^{-1}\hat{Z}_j' y_j = [Z_j'X(\hat{X}'X)^{-1}]^{-1}Z_j'X(X'X)^{-1}X'y_j,
$$

where $\hat{Z}_j = [\hat{Y}_j, X_j] = [Y_j, X_j]X(\hat{X}'X)^{-1}X'$. An interpretation of this formula is: 

i.) Regress $Y_j$ on $X$ to get a predicted $\hat{Y}_j$; 

ii.) estimate $\delta_j$ by OLS regression of $y_j$ on $\hat{Y}_j$ and $X_j$. 

• For an exactly identified equation, the two-stage least square estimator can be simplified to $\hat{\delta}_{j,2SLS} = (X'Z_j)^{-1}X'y_j$. This is so because both $X'Z_j$ is invertible. 

• It can be shown that $\hat{\delta}_{j,2SLS}$ is asymptotically the best instrumental variable estimator.

Denote $Z_1 = (Y_1, X_1)$. The structural equation is $y_{1.1} = Y_1\beta_{1.1} + X_1\gamma_{1.1} + u_{1.1} = Z_1\delta_{1.1} + u_{1.1}$. The 2SLS estimator can also be written as 

$$
\hat{\delta}_{1.1} = (Z_1'X(X'X)^{-1}X'Z_1)^{-1}Z_1'X(X'X)^{-1}X'y_{1.1},
$$

which is the minimizer of the following problem: 

$$
\min_{\delta} (y_{1.1} - Z_1\delta)'X(X'X)^{-1}X'(y_{1.1} - Z_1\delta).
$$

CONSISTENCY AND ASYMPTOTIC NORMALITY: 

The 2SLS estimator implies that 

$$
\hat{\delta}_{1.1} - \delta_{1.1} = (Z_1'X(X'X)^{-1}X'Z_1)^{-1}Z_1'X(X'X)^{-1}X'u_{1.1}.
$$

Let $J_1$ be the selection matrix such that $X_1 = XJ_1$. As 

$$
Z_1'X(X'X)^{-1}X'Z_1 = \left(\begin{array}{cc}
\hat{F}_1'X'X\hat{F}_1 & \hat{F}_1'X'X_1 \\
X_1'X\hat{F}_1 & X_1'X_1
\end{array}\right) = \left(\begin{array}{c}
\hat{F}_1' \\
J_1
\end{array}\right) X'X (\hat{F}_1 \hspace{.2cm} J_1),
$$

$$
\text{plim} \frac{1}{n} Z_1'X(X'X)^{-1}X'Z_1 = \left(\begin{array}{c}
\Pi_1 \\
J_1
\end{array}\right),
$$

which will be denoted by $A_1$. This limiting matrix will be nonsingular if $\text{plim} \frac{1}{n} X'X$ is a nonsingular matrix and $[\Pi_1, J_1]$ has full column rank. The full rank condition is the rank identification condition. Recall that $YB + X\Gamma = U$ implies $\Pi B + \Gamma = 0$. For the structural equation $y_{1.1} = Y_1\beta_{1.1} + X_1\gamma_{1.1} + u_{1.1}$, $\Pi (\begin{array}{c}
1 \\
-\beta_{1.1}
\end{array})' - (\begin{array}{c}
\gamma_{1.1} \\
0
\end{array})' = 0$. Partition $\Pi$ conformably as 

$$
\Pi = \left(\begin{array}{ccc}
\Pi_{G_1} & \Pi_{G_1m_1} & \Pi_{G_1m^*} \\
\Pi_{G^*1} & \Pi_{G^*m_1} & \Pi_{G^*m^*}
\end{array}\right),
$$

where $G^* = G - G_1$ and $m^* = m - (m_1 + 1)$. It follows that 

$$
\Pi_{G_1m_1}\beta_{1.1} + \gamma_{1.1} = \Pi_{G_1} (\begin{array}{c}
1 \\
-\beta_{1.1}
\end{array}) = 0.
$$

The rank condition for identification is $\text{rank}(\Pi_{G^*m_1}) = m_1$. Since $\Pi_{G^*m_1}\beta_{1.1} = \Pi_{G^*1}$, the rank condition is equivalent to $\text{rank}(\Pi_{G^*m_1}) = m_1$. Hence $[\Pi_1 \hspace{.2cm} J_1] = \left(\begin{array}{cc}
\Pi_{G_1m_1} & I_{G_1} \\
\Pi_{G^*m_1} & 0
\end{array}\right)$ has full column rank. Therefore $A_1$ is nonsingular. On the other hand, 

$$
\text{plim} \frac{1}{n} Z_1'X(X'X)^{-1}X'u_{1.1} = \text{plim} \frac{1}{n} \left(\begin{array}{c}
\hat{F}_1'X'u_{1.1} \\
X_1'X'u_{1.1}
\end{array}\right) = 0.
$$
The instrumental variable estimator is consistent and

\[
\sqrt{n}(\hat{\delta}_1 - \delta_1) = \left( \frac{1}{n} Z_1' X (X'X)^{-1} X' Z_1 \right) - \frac{1}{\sqrt{n}} Z_1' X (X'X)^{-1} X' u_1
\]

\[= A_1^{-1} \frac{1}{\sqrt{n}} Z_1' X (X'X)^{-1} X' u_1.\]

As \( \frac{1}{\sqrt{n}} Z_1' X (X'X)^{-1} X' u_1 = \frac{1}{\sqrt{n}} \left( \frac{\bar{P}' Y}{X_{1u_1}} \right) \left( \frac{\bar{J}_1}{\sqrt{n}} \right) X' u_1 \) and \( \frac{1}{\sqrt{n}} X' u_1 \xrightarrow{D} N(0, \sigma^2 \text{plim}_{1/n} X' X), \)

\[\sqrt{n}(\hat{\delta}_1 - \delta_1) \xrightarrow{D} N(0, \sigma^2 \text{plim}_{1/n} X' X).\]

**IV ESTIMATORS**

Consider the class of instrumental variables given by \( Y_1 = Z_1 \delta_1 + u_1 \) and \( Z_1 \) can be consistently estimated by the method of instrumental variables. Corresponding to the \( n \times (m_1 + G_1) \) matrix, suppose that \( P \) is a \( n \times (m_1 + G_1) \) matrix of instrumental variables for \( Z_1 \), which satisfies the conditions that \( \lim_{n \to \infty} \frac{1}{n} P' Z_1 \) exists and is a nonsingular matrix, \( \lim_{n \to \infty} \frac{1}{n} P' u_1 = 0 \), and \( \lim_{n \to \infty} \frac{1}{n} P' u_1' P = \sigma^2 \lim_{n \to \infty} \frac{1}{n} P' P. \) The IV estimator with \( P \) is

\[\hat{\delta}_{1,IV} = (P' Z_1)^{-1} P' Y_1.\]

The instrumental variable estimator is consistent and

\[\sqrt{n}(\hat{\delta}_{1,IV} - \delta_1) \xrightarrow{D} N \left( 0, \sigma^2 \text{plim}_{1/n} \left( \frac{P' Z_1}{n} \right)^{-1} \frac{P' P}{n} \left( \frac{Z_1' Z_1}{n} \right)^{-1} \right).\]

**OPTIMAL IV ESTIMATOR**

Consider the class of instrumental variables given by \( P = X A \) where \( A \) is a \( G \times (m_1 + G_1) \) matrix. Then

\[\sqrt{n}(\hat{\delta}_{1,IV} - \delta_1) \xrightarrow{D} N \left( 0, \sigma^2 \text{plim}_{1/n} \left( \frac{A' X' Z_1}{n} \right)^{-1} \frac{A' X' X A}{n} \left( \frac{Z_1' X A}{n} \right)^{-1} \right).\]

The 2SLS is a special IV estimator with \( A = (X'X)^{-1} X' Z_1 \). The Schwartz inequality implies that

\[Z_1' X (X'X)^{-1} X' Z_1 \geq Z_1' X A (A' X' X A)^{-1} A' X' Z_1, \]

and it becomes an equality for the particular \( A = (X'X)^{-1} X' Z_1 \). Hence the asymptotically optimal IV estimator (within the class of IV estimators with \( P = X A \) for some \( A \)) is \( P^* = X (X'X)^{-1} X' Z_1 \), which gives the 2SLS estimator.

**Nonlinear Simultaneous Equations Models and Nonlinear Two-Stage Least Squares**

Consider the estimation of the following structural equation in a system of nonlinear simultaneous equations:

\[y_i = f(Y_i, X_{1i}, \alpha_o) + u_i, \quad i = 1, \ldots, n,\]

where \( y_i \) is a scalar endogenous variable, \( Y_i \) is a vector of endogenous variables, \( X_{1i} \) is a vector of exogenous variables, \( \alpha_o \) is a \( k \)-dimensional vector of parameters, and \( \{u_i\} \) are i.i.d. disturbances with zero mean and variance \( \sigma^2 \). Let \( y, Y, X_1 \) and \( f \) denote the corresponding \( n \)-dimensional vectors or matrices. A nonlinear two-stage least squares NL2S estimator of \( \alpha_o \) is the value of \( \alpha \) that minimizes

\[S_{\alpha}(\alpha|W) = (y - f)' W (W' W)^{-1} W'(y - f),\]

where \( W = \left( \frac{\partial f}{\partial \alpha} \right) \).
where $W$ is some matrix of constants with rank at least equal to $k$. The variables of $W$ are instrumental variables.

The consistency of the NL2S estimator depends crucially on the asymptotic orthogonality condition that $\operatorname{plim}_{n \to \infty} W'u = 0$ (and some other regularity conditions). Let $f_\circ$ denote $f$ evaluated at $\alpha_\circ$. It follows that

$$\frac{1}{n} S_n = \frac{1}{n} (y - f)' W'(W'W)^{-1} W'(y - f)$$

$$= \frac{1}{n} (y - f_\circ + f_\circ - f)' W(W'W)^{-1} W'(y - f_\circ + f_\circ - f)$$

$$= \frac{1}{n} y'(W'W)^{-1} W'u + \frac{1}{n} (f_\circ - f)' W(W'W)^{-1} W'(f_\circ - f) + \frac{2}{n} u'(W'W)^{-1} W'(f_\circ - f).$$

The term $\frac{1}{n} u'(W'W)^{-1} W'u$ converges in probability to zero. The term $\frac{1}{n} (f_\circ - f)' W(W'W)^{-1} W'(f_\circ - f)$ has a unique minimum at $\alpha_0$ (assumption).

**Asymptotic Normality:**

We have

$$\frac{1}{\sqrt{n}} \frac{\partial^2 S_n}{\partial \alpha} \bigg|_{\alpha_\circ} = -\frac{2}{\sqrt{n}} \frac{\partial f'(\alpha_\circ)}{\partial \alpha} W'(W'W)^{-1} W'u$$

$$\xrightarrow{D} N(0, 4\sigma^2 \operatorname{plim} \frac{1}{n} \frac{\partial f'(\alpha_\circ)}{\partial \alpha} W(W'W)^{-1} W' \frac{\partial f'(\alpha_\circ)}{\partial \alpha}).$$

On the other hand, under the assumed conditions,

$$\operatorname{plim}_{n \to \infty} \frac{\partial^2 S_n}{\partial \alpha \partial \alpha'} \bigg|_{\alpha_\circ} = 2 \operatorname{plim} \frac{1}{n} \frac{\partial f(\alpha_\circ)}{\partial \alpha} W(W'W)^{-1} W' \frac{\partial f(\alpha_\circ)}{\partial \alpha'}.$$

The results follow from a Taylor series expansion that

$$\sqrt{n} (\hat{\alpha} - \alpha_\circ) \xrightarrow{D} N(0, \sigma^2 \left[ \operatorname{plim} \frac{1}{n} \frac{\partial f(\alpha_\circ)}{\partial \alpha} W(W'W)^{-1} W' \frac{\partial f(\alpha_\circ)}{\partial \alpha} \right]^{-1}).$$

The above NL2S method can be easily generalized to the estimation of the model $f(y_t, Y_t, X_{1t}, \alpha) = u_t$. The class of NL2S estimators is derived from

$$\min_{\alpha} f'(W'W)^{-1} W'(u_t - F(x_t, \beta_\circ)) = 0.$$

All the previous results are valid with apparent modifications

**The Formulation of GMM Estimators**

Let $\{x_t : t \geq 1\}$ be a stochastic process in $R^p$. Suppose that there exists a measurable function $f : R^p \times S \to R^q$ where $S$ is a parameter space of dimension $q$ such that $Ef(x, \beta)$ exists and is finite for all $\beta \in S$ and $Ef(x, \beta_0) = 0$. The latter condition is an orthogonality condition. Popular ways to obtain orthogonality conditions exploit the uncorrelatedness of the disturbances in an econometric model and instrumental variables. For an example, if $u_t = F(x_t, \beta_\circ)$ is the vector of disturbances of an econometric model which is uncorrelated with a vector of instrumental variables $z_t = G(x_t, \beta_\circ)$, then

$$Ef(u_t \otimes z_t) = 0.$$

For this example, the function $f$ is

$$f(x_t, \beta_\circ) = F(x_t, \beta_\circ) \otimes G(x_t, \beta_\circ).$$
Let \( \{x_t(\omega) : 1 \leq t \leq T\} \) be observable data of sample size \( T \). Denote \( f_t(\omega, \beta) = f(x_t(\omega), \beta) \). Let
\[
gr_T(\omega, \beta) = \frac{1}{T} \sum_{t=1}^{T} f_t(\omega, \beta)
\]
be a moment estimator of \( E(f_t(\omega, \beta)) \). Denote \( h_T(\omega, \beta) = a_T(\omega)g_T(\omega, \beta) \) where \( a_T(\omega) \) is a \( s \times r \)-dimensional random matrix with \( q \leq s \leq r \) such that \( a_T \) converges a.s. to a constant matrix \( a_0 \). A generalized method of moments (GMM) estimator is defined as
\[
\hat{\beta} = \arg\min_{\beta \in S} \|h_T(\omega, \beta)\|^2 = \arg\min_{\beta \in S} \langle g_T'(\omega, \beta) a_T(\omega) a_T(\omega) g_T(\omega, \beta) \rangle.
\]
For an example, consider the estimation of a nonlinear simultaneous equation model \( y_t = f(Y, x_t, \alpha) + u_t \). The Amemiya's NL2S estimation procedure is \( \min_{\alpha} (y - f)^T W (W' W)^{-1} W' (y - f) \), where \( W \) is some matrix of constants with a rank greater than the number of unknown parameters. The N2SL estimator is a GMM estimator with \( a_T(\omega) = (W' W / T)^{-1/2} \) and \( g_T(\omega, \beta) = W' (y - f) / T \). To investigate asymptotic properties of the GMM estimator, the stochastic process needs to satisfy some regularity properties.

Asymptotic Distribution of GMM

The GMM estimator \( \hat{\beta} \) satisfies the first order condition:
\[
\frac{\partial g_T'(\hat{\beta})}{\partial \beta} a_T g_T(\hat{\beta}) = 0.
\]
Let \( a_T^* \) denote \( \frac{\partial g_T'(\omega)}{\partial \beta} a_T^* a_T \). If \( a_T^* \) converges in probability to \( a_0^* \), the GMM estimator \( \hat{\beta} \) asymptotically satisfies the equation \( a_0^* E_f(x, \beta) = 0 \).

By a Taylor expansion,
\[
0 = a_T g_T(\hat{\beta}) = a_T^* g_T(\beta_0) + a_T^* \frac{\partial g_T'(\hat{\beta})}{\partial \beta} (\hat{\beta} - \beta_0),
\]
which implies that
\[
\sqrt{T}(\hat{\beta} - \beta_0) = -\left( a_T^* \frac{\partial g_T'(\hat{\beta})}{\partial \beta} \right)^{-1} \sqrt{T} a_T^* g_T(\beta_0)
\]
\[
= -\left( a_T^* \frac{\partial g_T'(\hat{\beta})}{\partial \beta} \right)^{-1} a_T^* \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(x_t, \beta_0).
\]

It follows that \( \sqrt{T}(\hat{\beta} - \beta_0) \) converges in distribution to a normally distributed random vector with mean zero and covariance matrix
\[
(a_0^* a_0^* a_0^* a_0^*)^{-1} a_0^* a_0^* a_0^* a_0^* (a_0^* a_0^* a_0^* a_0^*)^{-1},
\]
where \( S_w = \lim_{T \to \infty} \var( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(x_t, \beta_0) ) = \sum_{t=1}^{\infty} E(f(x_0, \beta_0) f(x_t, \beta_0)) \).

From this result, it is clear by the generalized Schwartz inequality that the optimal choice of \( a_0 \) is \( S_w^{-1/2} \). With the optimal \( a_0 \), it GMM method corresponds to
\[
\min_{\beta} g_T'(\beta) S_w^{-1} g_T(\beta),
\]
where \( S_w \) is a consistent estimate of \( S_w \). One can also show that \( T g_T'(\hat{\beta}) S_w^{-1} g_T(\hat{\beta}) \) is asymptotically chi-square distributed with \( r - q \) degrees of freedom. Since
\[
\sqrt{T}(\hat{\theta} - \beta_0) = -\left( \frac{\partial g_T'(\hat{\beta})}{\partial \beta} S_w^{-1} \frac{\partial g_T'(\hat{\beta})}{\partial \beta} \right)^{-1} \frac{\partial g_T'(\hat{\beta})}{\partial \beta} S_w^{-1} \sqrt{T} g_T(\beta_0),
\]
it follows
\[ \sqrt{Tg_T(\hat{\beta})} = \sqrt{Tg_T(\beta_o) + \frac{\partial g_T(\hat{\beta})}{\partial \beta} \sqrt{T}(\hat{\beta} - \beta_o)} \]
\[ = \left[I - \frac{\partial g_T(\hat{\beta})}{\partial \beta} \left( \frac{\partial g_T(\hat{\beta})}{\partial \beta} \hat{S}^{-1} \frac{\partial g_T(\hat{\beta})}{\partial \beta'} \right)^{-1} \frac{\partial g_T(\hat{\beta})}{\partial \beta} \right] \sqrt{Tg_T(\beta_o)}, \]
and
\[ S_w^{-1/2} \sqrt{Tg_T(\hat{\beta})} \]
\[ = \left[I - S_w^{-1/2} \frac{\partial g_T(\hat{\beta})}{\partial \beta} \left( \frac{\partial g_T(\hat{\beta})}{\partial \beta} \hat{S}^{-1} \frac{\partial g_T(\hat{\beta})}{\partial \beta'} \right)^{-1} \frac{\partial g_T(\hat{\beta})}{\partial \beta} \right] S_w^{-1/2} \sqrt{Tg_T(\beta_o)}. \]
The result holds because \( S_w^{-1/2} \sqrt{Tg_T(\beta_o)} \) converges in distribution to \( N(0, I) \) and the matrix in the bracket is idempotent with rank \( r - q \).