will fail to have full rank. For estimation of the restricted model, $\boldsymbol{W}$ must have at least k columns, while for running regression (11.08) it must have at least $2k$. If $\boldsymbol{W}$ has fewer than $2k$ columns, the test statistic will have fewer than k degrees of freedom and will actually be testing against a less general alternative than H_1. The obvious solution is effectively to double the number of instruments by using the matrix

$$\boldsymbol{W}^* \equiv \begin{bmatrix} \boldsymbol{W}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{W}_2 \end{bmatrix} \tag{11.09}$$

in place of $\boldsymbol{W}$ in the GNR (11.08). This allows the relationships between the endogenous regressors and the instruments to differ in the two parts of the sample, which seems quite reasonable. If one wants to use an LM test, that is, a test based on the explained sum of squares from regression (11.08), one must be careful to use $\boldsymbol{W}^*$ when one estimates the *restricted* model as well. However, as we discussed in Section 7.7, that is not necessary if one uses a $C(\alpha)$ test, that is, a pseudo-F test for $\boldsymbol{c} = \mathbf{0}$ in regression (11.08).

It is perhaps worth spelling out just how one should proceed if one wishes to test H_0 against H_1 when using IV estimation:

(i) Estimate the model H_0 using a suitable matrix $\boldsymbol{W}$ consisting of at least k, and preferably more than k, instruments, including all exogenous and predetermined variables in the regression function.

(ii) Create a new instrument matrix $\boldsymbol{W}^*$ as in (11.09). Then, to obtain the restricted SSR, run the GNR

$$\tilde{\boldsymbol{u}} = \boldsymbol{P}_{W^*}\tilde{\boldsymbol{X}}\boldsymbol{b} + \text{residuals}$$

over the entire sample, where $\tilde{\boldsymbol{u}}$ and $\tilde{\boldsymbol{X}}$ are evaluated at the IV estimates found in stage (i).

(iii) To obtain the unrestricted SSR, run the GNR

$$\tilde{\boldsymbol{u}}_j = \boldsymbol{P}_{W_j}\tilde{\boldsymbol{X}}_j\boldsymbol{b} + \text{residuals}$$

over each of the two subsamples separately and sum the two sums of squared residuals. Here $\tilde{\boldsymbol{u}}_j$, $\boldsymbol{W}_j$, and $\tilde{\boldsymbol{X}}_j$ denote the subvectors or submatrices of $\tilde{\boldsymbol{u}}$, $\boldsymbol{W}$, and $\tilde{\boldsymbol{X}}$ corresponding to the two subsamples.

(iv) Compute a $C(\alpha)$, or pseudo-F, test statistic based on the regression results obtained in (ii) and (iii), as described in Section 7.7.

An alternative procedure, which would be considerably more difficult in the nonlinear case, would be to estimate both the restricted and unrestricted models, using $\boldsymbol{W}^*$ for the instruments in both cases. For the unrestricted model, this would mean doing IV estimation for each part of the sample separately, using $\boldsymbol{W}_j$ as instruments for subsample j. Then one could calculate any of the test statistics based on restricted and unrestricted estimates that were discussed in Section 7.7.