Inference in Partially Identified Heteroskedastic Simultaneous Equations Models[1]

Helmut Lütkepohl
DIW Berlin and Freie Universität Berlin, Mohrenstr. 58, 10117 Berlin, Germany
email: hluetkepohl@diw.de

George Milunovich
Macquarie University, Sydney, NSW, 2109, Australia
email: george.milunovich@mq.edu.au

and

Minxian Yang
The University of New South Wales, Sydney, NSW, 2052, Australia
email: m.yang@unsw.edu.au

December 22, 2016

Abstract
Identification through heteroskedasticity in heteroskedastic simultaneous equations models (HSEMs) is considered. The possibility that heteroskedasticity identifies the structural parameters only partially is explicitly allowed for. The asymptotic properties of the identified parameters are derived. Moreover, tests for identification through heteroskedasticity are developed and their asymptotic distributions are derived. Monte Carlo simulations are used to explore the small sample properties of the asymptotically valid methods. Finally, the approach is applied to investigate the relation between the extent of economic openness and inflation.

JEL code: C30
Key words: Heteroskedasticity, simultaneous equations models, testing for identification, Davies' problem

[1]The research for this paper was partly carried out while the first author was a Bundesbank Professor at the Freie Universität Berlin. Financial support was provided by the Deutsche Forschungsgemeinschaft through SFB 649 "Economic Risk".

1 Introduction

Identifying the parameters in a simultaneous equations model (SEM) is typically a crucial step when employing SEMs for economic analysis. The identifying assumptions are often controversial and sometimes economic theory does not provide sufficiently many restrictions to fully identify all parameters. Econometricians have responded to this problem by developing methods for partially identified models (e.g., Phillips (1989), Choi and Phillips (1992)) or techniques for integrating extraneous information, e.g., in the form of extraneous instruments (e.g., Judge, Griffiths, Hill, Lütkepohl and Lee (1985)). The latter approach has the drawback that the instrumental variables (IV) may be weak which severely hampers inference. The weak instrument problem was pointed out by several authors (e.g., Staiger and Stock (1997), Dufour (2003)). One response has been to develop identification robust methods (e.g., Beaulieu, Dufour and Khalaf (2013), Doko Tchatoka and Dufour (2014)). Another option is to consider other types of information or data features such as heteroskedasticity or non-Gaussianity for identification (e.g., Lewbel (2012), Klein and Vella (2010), Farré, Klein and Vella (2013), Rigobon (2003), Lanne and Lütkepohl (2008)). In fact, Lewbel (2012) traces back related ideas to the work of Wright (1928).

In the present study we focus on heteroskedastic SEMs (HSEMs) which are identified through (conditional) heteroskedasticity while explicitly allowing for partial identification. In other words, only a subset of the structural parameters may be identified through heteroskedasticity while the remaining parameters may not be identified at all. A number of studies consider point inference in fully-identified HSEMs (see Klein and Vella (2010), Lewbel (2012), and Milunovich and Yang (2013) among others). However, in practice only a subset of the parameters in a HSEM can be identified through heteroskedasticity when an insufficient number of structural innovations exhibit heteroskedasticity. Therefore, in this article, we examine the partially-identified HSEM in the framework of Gaussian quasi maximum likelihood (QML), where only some of the structural equations are point identified. Within this context, a sequential procedure is proposed to estimate the identified equations. We find that the estimators of the identified parameters are consistent and asymptotically normal. Our simulation experiments indicate that the QML estimator performs well in finite samples and its root mean squared error decreases when the sample size increases.

Given that the question of which of the parameters are identified is central in our approach, we also develop tests for identification. More precisely, we consider tests for the heteroskedasticity rank of a HSEM which is a measure for the heterogeneity in the second moments of the structural errors and turns out to correspond to the number of structural equations which can be identified via heteroskedasticity. The tests are an instance of Davies' testing problem, where nuisance parameters are present only under the alternative hypothesis. We use the methods suggested by Hansen (1996) and Andrews and Ploberger (1994) to construct suitable tests for our purposes. In particular, the asymptotic null distributions of sup-LR and sup-LM test statistics for the hypotheses of interest in the present context are derived. In addition we propose a pragmatic residual-based test to sequentially determine the heteroskedasticity rank of HSEMs. Our simulation experiments show that the asymptotic null distributions of the tests are good approximations to their finite sample distributions, and that the tests exhibit powers which increase with the sample size.

Our approach is closely related to the methods used by Lanne and Saikkonen (2007) and Lütkepohl and Milunovich (2016) in a time series context for estimating a multivariate factor GARCH model and for testing identification in structural VAR-GARCH models, respectively. Although our approach is applicable to time series data, it is tilted towards cross-sectional data and does not cover the GARCH-type conditional heteroskedasticity considered by Lanne and Saikkonen (2007) and Lütkepohl and Milunovich (2016). Nevertheless, it does cover a wide range of conditional variance specifications. Our results complement the latter papers. We also present Monte Carlo evidence that our asymptotic results are a good indicator for the small sample properties of the estimators.

Our approach is statistics-based and does not depend on *a priori* economic restrictions on the parameter space, but instead relies on statistical properties of the model. Of course, economic information is still needed to interpret the equations and parameters properly. If the economic theory does not provide a fully identified model, the identifying restrictions from heteroskedasticity may complement the economic information. If the combined information from economic theory and heteroskedasticity is overidentifying, the restrictions can even be tested against the data. In particular, if the HSEM is fully identified through heteroskedasticity, any additional restrictions from economic considerations can be tested with statistical tools. Specifically, competing economic theories can be tested against the data if identification is provided through heteroskedasticity. We will illustrate the usefulness and significance of our approach by reconsidering the problem of whether openness of an economy has an impact on inflation. This issue has been studied by Romer (1993) who argues that openness reduces inflation. Using our approach we can resolve endogeneity problems in his study.

The structure of our study is as follows. In the next section we present the model setup and in Section 3 we discuss estimation procedures and asymptotic properties of the estimators. Testing the heteroskedasticity rank is considered in Section 4 and small sample Monte Carlo results are presented in Section 5. Section 6 considers the empirical illustration and Section 7 concludes. All proofs are collected in the Appendix.

2 The Model

We consider the structural-form simultaneous equation model

$$Ay_i = Cx_i + \varepsilon_i, \qquad i = 1, \dots, n, \tag{1}$$

where i is the observation index, n is the number of observations, y_i and x_i are K and K_x dimensional observable vectors of endogenous and exogenous variables respectively, A (invertible) and C are coefficient matrices of dimensions $K \times K$ and $K \times K_x$ respectively. The K-dimensional structural error vector ε_i is assumed to have the following properties:

$$\mathbb{E}(\varepsilon_i | W_i) = 0, \qquad \text{var}(\varepsilon_i | W_i) = H_i, \tag{2}$$

where W_i is the set of all observable exogenous or predetermined variables (including x_i),

$$H_i = \begin{bmatrix} \Lambda_i & 0 \\ 0 & I_{K-r} \end{bmatrix}, \qquad \Lambda_i = \text{diag}[\sigma^2_{1,i}, \dots, \sigma^2_{r,i}], \qquad \mathbb{E}(\Lambda_i) = I_r,$$

$\sigma^2_{k,i} = \exp\{F_k(z_i, \beta_k)\}$ with $z_i \in W_i$. The function $F_k(z_i, \beta_k)$ is twice continuously differentiable with respect to β_k for $k = 1, \dots, r$ and $0 \leq r \leq K$. Note that $H_i = \Lambda_i$ when $r = K$ and $H_i = I_K$ when $r = 0$.

Our model allows for the possibility that $K - r$ of the structural errors are homoskedastic and r errors are conditionally heteroskedastic. The structural form is set up such that the first r errors are (conditionally) heteroskedastic while the last $K - r$ errors are (conditionally) homoskedastic. The standardization of the conditional variances of these last errors to be one does not entail a loss of generality because we do not impose restrictions on the structural coefficient matrix A. In particular, the diagonal of A is not normalized to be a unit diagonal. Thus, the K equations in (1) may not directly provide economically meaningful interpretations. In practice, however, there may be restrictions or at least some features of the structural parameters that make the equations structurally interpretable. We explicitly do not impose such restrictions at this point because we are interested in studying to what extent identification comes from conditional heteroskedasticity and how much is needed in addition from other sources. If all parameters turn out to be identified through conditional heteroskedasticity, then any other identification restrictions become overidentifying and can be tested against the data. This is an important advantage of our approach, provided that there is enough identifying information from the covariance structure.

Our main interest is in the cases with $1 \leq r < K-1$, where only a subset of the parameters in (1) and (2) is point identified. We assume that the conditional variances in Λ_i are linearly independent, and will call r the heteroskedasticity rank. The structural error ε_i can then be written as $\varepsilon_i = H_i^{1/2}\eta_i$, where the standardized error satisfies $\mathbb{E}(\eta_i|W_i) = 0$ and $\text{var}(\eta_i|W_i) = I_K$. Note that in this setting the unconditional variance of ε_i is normalized as the identity matrix, i.e., $\text{var}(\varepsilon_i) = I_K$.

The reduced-form for model (1) is given by

$$y_i = Dx_i + u_i, \qquad u_i = B\varepsilon_i, \qquad D = BC, \qquad B = A^{-1}. \tag{3}$$

The unconditional variance matrix of the reduced-form error is $\text{var}(u_i) = \Omega = BB'$. The parameters of the reduced-form model, D and Ω, can be consistently estimated by ordinary least squares (OLS) applied to each equation separately.

Following Lanne and Saikkonen (2007) and Lütkepohl and Milunovich (2016), we use the partitioning $B = [B_1, B_2]$, where B_1 and B_2 are respectively the first r columns and the last $K-r$ columns of B. Conformably, $A' = [A_1', A_2']$. As $u_i = B_1\varepsilon_i^{(1:r)} + B_2\varepsilon_i^{(r+1:K)}$, where $\varepsilon_i^{(1:r)}$ and $\varepsilon_i^{(r+1:K)}$ denote respectively the first r and the last $K-r$ elements of ε_i, the conditional variance of u_i is

$$\Omega_i = \text{var}(u_i|W_i) = B_1\Lambda_i B_1' + B_2B_2' = \Omega + B_1(\Lambda_i - I_r)B_1'.$$

For Gaussian quasi maximum likelihood (QML) estimation, the conditional probability density function (pdf) for the reduced-form error $u_i = y_i - Dx_i$ is given by

$$\text{pdf}(u_i|W_i) = (2\pi)^{-\frac{K}{2}}\det(\Omega_i)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}u_i'\Omega_i^{-1}u_i\right\}.$$

Because $\det(\Omega_i) = \det(\Omega)\det(H_i)$ and $\Omega_i^{-1} = A'H_i^{-1}A = \Omega^{-1} + A_1'(\Lambda_i^{-1} - I_r)A_1$, the conditional pdf becomes

$$\begin{aligned}\text{pdf}(u_i|W_i) &= (2\pi)^{-\frac{K}{2}}\det(\Omega)^{-\frac{1}{2}}\det(\Lambda_i)^{-\frac{1}{2}} \\ &\quad \times \exp\left\{-\frac{1}{2}u_i'\Omega^{-1}u_i\right\}\exp\left\{-\frac{1}{2}u_i'A_1'(\Lambda_i^{-1}-I_r)A_1u_i\right\}, \qquad (4)\end{aligned}$$

which is also valid for $r = K$ with A_1 being equal to A. If $r \geq K-1$ and the conditional variances are not proportional, then A is fully identified (see Milunovich and Yang (2013)). However, for $r < K-1$, A_2 is unidentified as it is absorbed into Ω. Thus, if $r < K-1$, the structural form is only partially identified. More precisely, only the parameters in the first r quations are identified by heteroskedaticity if the heteroskedasticity rank is less than $K-1$.

3 Estimation

The log quasi likelihood (apart from the constant $\frac{1}{2}nK\ln(2\pi)$) is given by

$$\mathcal{L}_n = -\frac{nK}{2}\ln|\Omega| - \frac{1}{2}\sum_{i=1}^{n}\left[u_i'\Omega^{-1}u_i + \ln|\Lambda_i| + u_i'A_1'(\Lambda_i^{-1}-I_r)A_1u_i\right], \qquad (5)$$

where ln denotes the natural logarithm. At this point we assume that D is known or u_i is observable. We will show later on that our results hold when u_i is replaced by the reduced-form OLS residual $\hat{u}_i$. For any given $(A_1, \beta_1, \ldots, \beta_r)$, where β_k is the parameter vector in $\sigma_{k,i}^2$, (5) is maximized by $\hat{\Omega} = n^{-1}\sum_{i=1}^{n}u_iu_i'$. Substituting $\hat{\Omega}$ into (5) yields

$$\begin{aligned}\mathcal{L}_n &= -\frac{nK}{2}(\ln|\hat{\Omega}|+1) + \frac{n}{2}\sum_{k=1}^{r}\ell_{k,n}, \qquad (6)\\ \ell_{k,n} &= -\frac{1}{n}\sum_{i=1}^{n}\left[\ln(\sigma_{k,i}^2) + a_k'u_iu_i'a_k(\sigma_{k,i}^{-2}-1)\right], \qquad k=1,\ldots,r,\end{aligned}$$

where a_k' is the k^{th} row of A_1. The estimators of $(A_1, \beta_1, \ldots, \beta_r)$ are the maximizers of (6), subject to the restriction $A_1\hat{\Omega}A_1' = I_r$. With this restriction, (6) is also valid for the case where $r = K$. The estimators of $(A_1, \beta_1, \ldots, \beta_r)$ are the maximizers of (6). They may be obtained by maximizing $\ell_{k,n}$ for $k = 1, \ldots, r$ sequentially. In a time series context, Lanne and Saikkonen (2007) sequentially maximize $\ell_{k,n}$ to obtain starting values for the overall maximization of their quasi likelihoods. Our setup differs from their setup in that the parameters in a_k are variation free from those in $\sigma_{k,i}^2$. Thus, sequential maximization of the $\ell_{k,n}$ results in the overall maximum.

We now describe the estimation procedure and show that the estimators obtained are consistent for the columns of A_1' and $\beta = [\beta_1', \ldots, \beta_r']'$ at the true parameter point.

3.1 Estimation Procedure

First, we estimate (a_1, β_1) by maximizing $\ell_{1,n}$. For a given $\sigma^2_{1,i}$ (or β_1), the quadratic form $a_1' n^{-1} \sum_{i=1}^n u_i u_i' (\sigma_{1,i}^{-2} - 1) a_1$, with the restriction $a_1' \hat{\Omega} a_1 = 1$, is minimized by the eigenvector $\hat{a}_1$ associated with the smallest generalized eigenvalue $\hat{\mu}_1$ in

$$(\Psi_{1,n} - \mu_1 \hat{\Omega}) a_1 = 0, \tag{7}$$

where $\Psi_{1,n} = n^{-1} \sum_{i=1}^n u_i u_i' (\sigma_{1,i}^{-2} - 1)$ and $a_1' \hat{\Omega} a_1 = 1$. Then, the concentrated objective function

$$\ell_{1,n}(\beta_1) = -\frac{1}{n} \sum_{i=1}^n \ln(\sigma^2_{1,i}) - \hat{\mu}_1,$$

is maximized, where both $\sigma_{1,i}$ and $\hat{\mu}_1$ are functions of β_1. The estimator of β_1 is $\hat{\beta}_1 = \arg\max_{\beta_1} \ell_{1,n}(\beta_1)$. The estimator of a_1, denoted as $\hat{a}_1$, is the eigenvector obtained from (7), where all unknown quantities are evaluated at $\beta_1 = \hat{\beta}_1$. For a given β_1, $\hat{a}_1$ is completely determined by (7). Hence, the evaluation of $\ell_{1,n}(\beta)$ may be carried out in two steps: (a) the smallest eigenvalue $\hat{\mu}_1$ from (7) is computed for a given β_1; (b) $\ell_{1,n}(\beta_1) = -n^{-1} \sum_{i=1}^n \ln(\sigma_{1,i}) - \hat{\mu}_1$ is computed. The numerical maximization is done over the space of β_1 only.

Once $(\hat{a}_1, \hat{\beta}_1)$ are obtained, the estimators of (a_2, β_2) are the maximizers of $\ell_{2,n}$, subject to the restrictions $a_1' \hat{\Omega} a_2 = 0$ and $a_2' \hat{\Omega} a_2 = 1$. Let the matrix $[\hat{a}_1, Q_2]$ contain all the eigenvectors of (7) evaluated at $\hat{\beta}_1$, where Q_2 is a $K \times (K-1)$ matrix satisfying $\hat{a}_1' \hat{\Omega} Q_2 = 0$ and $Q_2' \hat{\Omega} Q_2 = I_{K-1}$. To implement the first restriction, a_2 is written as $a_2 = Q_2 \rho_2$, where ρ_2 is a $(K-1)$-dimensional vector satisfying $\rho_2' \rho_2 = 1$. The objective function can then be expressed as

$$\ell_{2,n} = -\frac{1}{n} \sum_{i=1}^n \Big[\ln(\sigma^2_{2,i}) + \rho_2' Q_2' u_i u_i' Q_2 \rho_2 (\sigma_{2,i}^{-2} - 1) \Big].$$

For given $\sigma^2_{2,i}$ (or β_2), the quadratic form $\rho_2' Q_2' u_i u_i' Q_2 \rho_2 (\sigma_{2,i}^{-2} - 1)$, subject to $\rho_2' \rho_2 = 1$, is minimized by the eigenvector $\hat{\rho}_2$ associated with the smallest eigenvalue $\hat{\mu}_2$ in

$$(Q_2' \Psi_{2,n} Q_2 - \mu_2 I_{K-1}) \rho_2 = 0, \tag{8}$$

where $\Psi_{2,n} = \frac{1}{n} \sum_{i=1}^n u_i u_i' (\sigma_{2,i}^{-2} - 1)$. Then the concentrated objective function

$$\ell_{2,n}(\beta_2) = -\frac{1}{n} \sum_{i=1}^n \ln(\sigma^2_{2,i}) - \hat{\mu}_2$$

is maximized to obtain the estimator $\hat{\beta}_2 = \arg\max_{\beta_2} \ell_{2,n}(\beta_2)$. The restriction $a_2 = Q_2\rho_2$ ensures that $\hat{\mu}_2 > \hat{\mu}_1$ at the maximizer $\hat{\beta}_2$. Let the matrix $[\hat{\rho}_2, R_3]$ contain all the eigenvectors of (8) evaluated at $\hat{\beta}_2$ and let $Q_3 = Q_2R_3$. Then, a_3 should be estimated from the space spanned by Q_3, i.e., $a_3 = Q_3\rho_3$. This way, we can further estimate (a_k, β_k) for $k = 3, \ldots, r$. The last column of A_1' is estimated by $\hat{a}_r = Q_r\hat{\rho}_r$. Let the matrix $[\hat{\rho}_r, R_{r+1}]$ contain all the eigenvectors of the r^{th} eigen equation evaluated at $\hat{\beta}_r$. It follows that $\hat{A}_2' = Q_{r+1} = Q_rR_{r+1}$ estimates the space spanned by A_2'. In the case of $r = K-1$, the model is fully-identified and $[\hat{a}_1, \ldots, \hat{a}_r, \hat{A}_2]$ estimates all columns in A'. We reiterate that this sequential procedure is equivalent to simultaneously maximizing (6) over $(A_1, \beta_1, \ldots, \beta_r)$, subject to the restrictions $A_1\hat{\Omega}A_1' = I_r$. We summarize the estimation procedure as follows.

(a) Set $k = 1$ and $Q_1 = \hat{\Omega}^{-1/2}$, which is the upper triangular Cholesky factor satisfying $Q_1'Q_1 = \hat{\Omega}^{-1}$.

(b) Find $\hat{\beta}_k = \arg\max_{\beta_k} \left\{ -\frac{1}{n}\sum_{i=1}^n \ln(\sigma_{k,i}^2) - \hat{\mu}_k \right\}$ over the space of β_k, where $\hat{\mu}_k$ is the smallest eigenvalue in

$$(Q_k'\Psi_{k,n}Q_k - \mu I_{K-k+1})\,\rho = 0 \qquad \text{with} \qquad \Psi_{k,n} = \frac{1}{n}\sum_{i=1}^n u_iu_i'(\sigma_{k,i}^{-2} - 1).$$

Find $\hat{a}_k = Q_k\hat{\rho}_k$, where $\hat{\rho}_k$ is the eigenvector associated with the smallest eigenvalue $\hat{\mu}_k$ evaluated at $\hat{\beta}_k$. Set $Q_{k+1} = Q_kR_{k+1}$, where the columns of R_{k+1} are eigenvectors associated with the $K-k$ largest eigenvalues evaluated at $\hat{\beta}_k$.

(c) If $k < r$, set $k = k+1$ and go to (b). Otherwise, set $\hat{A}_2' = Q_{k+1}$ and stop.

We note that this procedure is valid for all r, $1 \leq r \leq K$.

3.2 Consistency

To examine the asymptotic behavior of the QML estimator of $\theta = [\text{vec}(A_1')', \beta_1', \ldots, \beta_r']'$, we list a set of assumptions below. We define $f_{k,i} = \partial F_k(z_i, \beta_k)/\partial\beta_k$ and we use an innermost subscript 0 to indicate that the associated quantities are evaluated at the true parameter point θ_0. For instance, $\sigma_{0k,i}^2$ is $\sigma_{k,i}^2$ evaluated at θ_0, and a_{0k} is the kth row of A_{01}. Furthermore, we define $\mathcal{F}_i$ to be the sigma-field generated by $\{(W_j, u_j) : j = 1, \ldots, i\}$.

Assumption A

A1 The observable arrays $\{y_i, W_i\}_{i=1}^n$ are drawn from the data generating process (DGP) specified in equations (1)-(2).

A2 The standardised errors $\eta_i = H_i^{-1/2}\varepsilon_i$ are independent draws from a distribution with mean zero and variance I_K. The elements of W_i have finite second moments. The matrix $\mathbb{E}(x_i x_i')$ is of full rank.

A3 The arrays $\{y_i, W_i\}_{i=1}^n$ are independent across i for cross-sectional data, or are strictly stationary and ergodic for time series data (or for panel data in the time dimension).

A4 In a neighborhood of the true parameter point θ_0, $\mathcal{N}_{\theta_0}$, the log conditional variance $\ln \sigma_{k,i}^2 = F_k(z_i, \beta_k)$ is bounded by a function $g(\cdot)$ such that $\sup_{\mathcal{N}_{\theta_0}} |F_k(z_i, \beta_k)| \leq g(z_i)$ for $k \in \{1, \ldots, r\}$. Further, $\mathbb{E}g(z_i)$, $\mathbb{E}[u_i' u_i \exp\{g(z_i)\}]$, and $\mathbb{E}[x_i' x_i \exp\{g(z_i)\}]$ are finite.

A5 Let v_i be a K_v-dimensional $\mathcal{F}_i$-measurable random vector. For any K_v-dimensional constant vector $c \neq 0$,
(i) $\mathbb{E}(c' v_i | \mathcal{F}_{i-1}) = 0$;
(ii) $\mathbb{E}[(\max_{i \leq n} |c' v_i|)^2]/n$ is finite uniformly over n;
(iii) $\max_{i \leq n} |c' v_i|/n^{1/2} \xrightarrow{p} 0$;
(iv) $\sum_{i=1}^n (c' v_i)^2/n$ converges in probability to a positive constant.

Here, A1 simply asserts that the model considered is the data generating process. A2 and A3 are needed for applying the weak law of large numbers (WLLN) to the second moments of the data. A3 and A4 are technical conditions that enable us to apply a uniform WLLN to $\ell_{k,n}$. A5 spells out the requirements for applying the central limit theorem (CLT) of McLeish (1974) to the vector v_i via the Cramér-Wold device (Cramér and Wold (1936)). In our context, v_i will be either $\text{vec}(u_i x_i')$ or the score of the log quasi likelihood, which involves quantities $(\sigma_{0k,i}^{-2} - 1)u_i u_i' a_{0k}$ and $(1 - \sigma_{0k,i}^{-2} a_{0k}' u_i u_i' a_{0k}) f_{0k,i}$. In A5, with v_i being either $\text{vec}(u_i x_i')$ or the score, condition (i) holds for cross-sectional data when the data are random draws from a population and (i) holds for time series data quite generally. The following proposition states the consistency of our estimators under Assumption A.

Proposition 1. *If Assumptions A1-A4 hold and A5 holds for* $v_i \equiv \text{vec}(u_i x_i')$ *and for* $v_i \equiv \text{vec}(u_i x_i')\sigma_{k,i}^{-2}$ *in* $\mathcal{N}_{\theta_0}$*, the estimators* $(\hat{a}_k, \hat{\beta}_k)$ *for* $k = 1, \ldots, r$ *obtained from the sequential procedure described in this section are consistent in the following sense:*

$$\hat{\beta}_k \xrightarrow{p} \beta_{0l_k}, \qquad \hat{a}_k \xrightarrow{p} \pm a_{0l_k}, \qquad k = 1, \ldots, r,$$

where (a_{0l_k}, β_{0l_k}) *are the true parameters associated with the* l_k^{th} *equation in model (1), and* $l_k \in \{1, \ldots, r\}$*. Further, the order of* l_k *is determined by* $\mathbb{E}\ln(\sigma_{0l_k,i}^2)$*, i.e.,* $\mathbb{E}\ln(\sigma_{0l_1,i}^2) \leq \mathbb{E}\ln(\sigma_{0l_2,i}^2) \leq \cdots \leq \mathbb{E}\ln(\sigma_{0l_r,i}^2)$*. The above results also hold when* u_i *is replaced by the reduced-form residual* $\hat{u}_i$*.* □

The proposition is proven in the Appendix. There are two notable features in Proposition 1. First, $\hat{a}_k$ is consistent for a row of A_{01} up to the scale ± 1. This feature reflects the fact that the model described by (1) and (2) can only define each row of A_{01} up to the scale ± 1, as multiplying a row of A_{01} by -1 will lead to an observationally equivalent system. Second, the estimators $(\hat{a}_k, \hat{\beta}_k)$ are consistent for the true parameters in the equation with the k^{th} smallest mean of the log conditional variance ($k \in \{1, \ldots, r\}$). Put differently, $\hat{a}_k$ is a consistent estimator of the row in A_0 with the k^{th} smallest mean log conditional error variance.

3.3 Asymptotic Distribution

In what follows, without loss of generality, we assume that

$$\mathbb{E}\ln(\sigma_{01,i}^2) \leq \mathbb{E}\ln(\sigma_{02,i}^2) \leq \cdots \leq \mathbb{E}\ln(\sigma_{0r,i}^2),$$

i.e., (1) is arranged such that the mean log conditional variance of the first equation is the smallest and that of the r^{th} equation is the largest. As mentioned before, there is an indeterminancy of A_{01}' in the model (multiplying -1 to a column of A_{01}' gives rise to an observationally-equivalent system). To avoid this indeterminacy, without loss of generality, we define the columns of A_{01}' as the probability limit of the QML estimator $[\hat{a}_1, \ldots, \hat{a}_r]$. Let $\beta = [\beta_1', \ldots, \beta_r']'$ with dimension K_β and $\theta = [\text{vec}(A_1')', \beta_1', \ldots, \beta_r']'$ with dimension $K_\theta = rK + K_\beta$. We write the Lagrangian for the maximization of the log likelihood in (6) as

$$\begin{aligned} L_n &= -\frac{1}{2n}\sum_{i=1}^{n}\big(\ln|\Lambda_i| + u_i'A_1'(\Lambda_i^{-1} - I_r)A_1u_i\big) + \frac{1}{2}\mu'\text{vech}(A_1\hat{\Omega}A_1' - I_r), \\ &= \frac{1}{n}\mathcal{L}_n(\theta) + \mu'\phi(\theta), \end{aligned} \tag{9}$$

where $\phi(\theta) = \frac{1}{2}\text{vech}(A_1\hat{\Omega}A_1' - I_r)$, $\mu' = [\mu_{11}, \ldots, \mu_{r1}, \mu_{22}, \ldots, \mu_{r2}, \ldots, \mu_{rr}]$ is the vector of Lagrangian multipliers, which can be viewed as the vectorization of a symmetric $(r \times r)$ matrix.

To find the derivatives of L_n and the Jacobian of $\phi(\theta)$, we note that

$$\begin{aligned}
u_i'A_1'(\Lambda_i^{-1} - I_r)A_1u_i &= \text{vec}(A_1')'\big((\Lambda_i^{-1} - I_r) \otimes u_iu_i'\big)\text{vec}(A_1'),\\
\mu'\text{vech}(A_1\hat{\Omega}A_1') &= \mu'\mathcal{D}_r^{+}\text{vec}(A_1\hat{\Omega}A_1') = \text{vec}(\mathcal{M})'\text{vec}(A_1\hat{\Omega}A_1')\\
&= \text{tr}(\mathcal{M}A_1\hat{\Omega}A_1') = \text{vec}(A_1')'(\mathcal{M} \otimes \hat{\Omega})\text{vec}(A_1'),\\
(\mathcal{M} \otimes \hat{\Omega})\text{vec}(A_1') &= (I_r \otimes \hat{\Omega}A_1')\text{vec}(\mathcal{M}) = (I_r \otimes \hat{\Omega}A_1')\mathcal{D}_r^{+\prime}\mu,
\end{aligned}$$

where $\mathcal{D}_r^{+} = (\mathcal{D}_r'\mathcal{D}_r)^{-1}\mathcal{D}_r'$ and $\mathcal{D}_r$ is the $r^2 \times (r+1)r/2$ duplication matrix, defined such that $\mathcal{D}_r\text{vech}(\Psi) = \text{vec}(\Psi)$ for any $(r \times r)$ symmetric matrix Ψ, and $\mathcal{M} = [m_{ij}]$ is a $(r \times r)$ symmetric matrix with entries being $m_{ij} = \mu_{ij}$ if $i = j$ and $m_{ij} = .5\mu_{ij}$ if $i > j$. The first derivatives of L_n are given by

$$\frac{\partial L_n}{\partial \text{vec}(A_1')} = -\frac{1}{n}\sum_{i=1}^{n}\Big((\Lambda_i^{-1} - I_r) \otimes u_iu_i'\Big)\text{vec}(A_1') + (I_r \otimes \hat{\Omega}A_1')\mathcal{D}_r^{+\prime}\mu,$$

$$\frac{\partial L_n}{\partial \beta} = -\frac{1}{2n}\sum_{i=1}^{n}\begin{bmatrix}(1 - \sigma_{1,i}^{-2}a_1'u_iu_i'a_1)f_{1,i}\\ \vdots \\ (1 - \sigma_{r,i}^{-2}a_r'u_iu_i'a_r)f_{r,i}\end{bmatrix}.$$

The second derivatives of L_n are given by

$$\frac{\partial^2 L_n}{\partial \text{vec}(A_1')\partial \text{vec}(A_1')'} = -\frac{1}{n}\sum_{i=1}^{n}\Big((\Lambda_i^{-1} - I_r) \otimes u_iu_i'\Big) + (\mathcal{M} \otimes \hat{\Omega}),$$

$$\frac{\partial^2 L_n}{\partial \text{vec}(A_1')\partial \beta'} = \frac{1}{n}\sum_{i=1}^{n}(\Lambda_i^{-1} \otimes u_iu_i')\begin{bmatrix} a_1f_{1,i}' & \cdots & 0\\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_rf_{r,i}'\end{bmatrix},$$

$$\frac{\partial^2 L_n}{\partial \beta\partial \beta'} = -\frac{1}{2n}\sum_{i=1}^{n}\text{diag}\left\{\begin{bmatrix}(1 - \sigma_{1,i}^{-2}a_1'u_iu_i'a_1)\partial f_{1,i}/\partial \beta_1'\\ \vdots \\ (1 - \sigma_{r,i}^{-2}a_r'u_iu_i'a_r)\partial f_{r,i}/\partial \beta_r'\end{bmatrix} + \begin{bmatrix}(\sigma_{1,i}^{-2}a_1'u_iu_i'a_1)f_{1,i}f_{1,i}'\\ \vdots \\ (\sigma_{r,i}^{-2}a_r'u_iu_i'a_r)f_{r,i}f_{r,i}'\end{bmatrix}\right\}.$$

We denote the negative score by $S_n(\theta) = -n^{-1}\nabla_\theta \mathcal{L}_n$, the negative Hessian by $J_n(\theta) =$

$-n^{-1}\partial^2\mathcal{L}_n/\partial\theta\partial\theta'$, and $J_0 = \mathbb{E}J_n(\theta_0)$. It can be verified that

$$
\begin{aligned}
J_0 &= \begin{bmatrix} J_{0,11} & J_{0,12} \\ J_{0,12}' & J_{0,22} \end{bmatrix} \\
&= \mathbb{E}\begin{bmatrix} (\Lambda_{0i}^{-1} - I_r) \otimes B_0 H_{0i} B_0' & -(\Lambda_{0i}^{-1} \otimes B_0 H_{0i} B_0')\mathrm{diag}(a_{01} f_{01,i}', \dots, a_{0r} f_{0r,i}') \\ J_{0,12}' & \frac{1}{2}\mathrm{diag}(f_{01,i} f_{01,i}', \dots, f_{0r,i} f_{0r,i}') \end{bmatrix}. \quad (10)
\end{aligned}
$$

When $J_{0,22}$ is invertible, J_0 is invertible if and only if $J_{0,11} - J_{0,12}J_{0,22}^{-1}J_{0,12}'$ is invertible. The latter quantity can be expressed as

$$
J_{0,11} - J_{0,12}J_{0,22}^{-1}J_{0,12}' = (I_r \otimes B_0)\Big(\mathbb{E}\big[(\Lambda_{0i}^{-1} - I_r) \otimes H_{0i}\big] - 2\mathrm{diag}(G_1, \dots, G_r)\Big)(I_r \otimes B_0'),
$$

where $G_k = e_K^k e_K^{k\prime} \mathbb{E}(f_{0k,i})'[\mathbb{E}(f_{0k,i} f_{0k,i}')]^{-1}\mathbb{E}(f_{0k,i})$ and e_K^k is the k^{th} column of I_K for $k \in \{1, \dots, r\}$. The k^{th} (diagonal) block of $\mathbb{E}\big[(\Lambda_{0i}^{-1} - I_r) \otimes H_{0i}\big]$ is

$$
\mathbb{E}\big[(\sigma_{0k,i}^{-2} - 1)\mathrm{diag}(\sigma_{01,i}^2, \dots, \sigma_{0r,i}^2, 1, \dots, 1)\big], \qquad k \in \{1, \dots, r\},
$$

which is a diagonal matrix with one zero at the k^{th} diagonal position. Therefore, when $J_{0,22}$ is invertible and $\mathbb{E}(f_{0k,i}) \neq 0$ for every k, J_0 is invertible. However, J_0 is not positive definite. The negative score at θ_0 is given by

$$
S_n(\theta_0) = \begin{bmatrix} -\dfrac{\partial\mathcal{L}_n}{\partial\mathrm{vec}(A_1')} \\ -\dfrac{\partial\mathcal{L}_n}{\partial\beta} \end{bmatrix}_{\theta_0} = \frac{1}{n}\sum_{i=1}^n s_i(\theta_0) = \frac{1}{n}\sum_{i=1}^n \begin{bmatrix} \Big((\Lambda_{0i}^{-1} - I_r) \otimes u_i u_i'\Big)\mathrm{vec}(A_{01}') \\ \frac{1}{2}(1 - \sigma_{01,i}^{-2} a_{01}' u_i u_i' a_{01}) f_{01,i} \\ \vdots \\ \frac{1}{2}(1 - \sigma_{0r,i}^{-2} a_{0r}' u_i u_i' a_{0r}) f_{0r,i} \end{bmatrix},
$$

where $s_i(\theta_0)$ is defined as

$$
s_i(\theta_0) = \begin{bmatrix} \Big((\Lambda_{0i}^{-1} - I_r) \otimes u_i u_i'\Big)\mathrm{vec}(A_{01}') \\ \frac{1}{2}(1 - \sigma_{01,i}^{-2} a_{01}' u_i u_i' a_{01}) f_{01,i} \\ \vdots \\ \frac{1}{2}(1 - \sigma_{0r,i}^{-2} a_{0r}' u_i u_i' a_{0r}) f_{0r,i} \end{bmatrix}.
$$

It can be verified that $\mathbb{E}\big(s_i(\theta_0)\big) = 0$.

There are $K_\phi = r(r+1)/2$ restrictions in $\phi(\theta)$. We write $\phi(\theta)' = [\phi_1(\theta), \dots, \phi_{K_\phi}(\theta)]$ and $\Phi(\theta) = \nabla_\theta\phi(\theta)' = [\nabla_\theta\phi_1(\theta), \dots, \nabla_\theta\phi_{K_\phi}(\theta)]$, where $\nabla_\theta \equiv \frac{\partial}{\partial\theta}$. From the first derivatives of the Lagrangian L_n, the Jacobian of the constraints is seen to be

$$
\Phi(\theta) = \nabla_\theta\phi(\theta)' = \begin{bmatrix} (I_r \otimes \hat{\Omega} A_1')\mathcal{D}_r^{+\prime} \\ 0 \end{bmatrix},
$$

where 0 denotes a $K_\beta \times K_\phi$ zero matrix. The Taylor expansion of $\phi(\hat{\theta}) = 0$ at θ_0 gives

$$0 = \phi(\hat{\theta}) = \phi(\theta_0) + \Phi(\bar{\theta})'(\hat{\theta} - \theta_0) = \Phi(\bar{\theta})'(\hat{\theta} - \theta_0),$$

where $\bar{\theta}$ is a point between $\hat{\theta}$ and θ_0. This implies, as $n \to \infty$, $\Phi(\theta_0)'\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{p} 0$, i.e., the asymptotic variance of $\sqrt{n}(\hat{\theta} - \theta_0)$ is singular. We describe the asymptotic distribution of $\hat{\theta}$ in the following proposition.

Proposition 2. *Suppose that the assumptions in Proposition 1 hold. Assume further that A5 holds for $v_i \equiv s_i(\theta_0)$ and that $\mathbb{E}(f_{0k,i} f'_{0k,i})$ is of full rank for $k = 1, \ldots, r$. Then,*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma_\theta),$$

where the asymptotic covariance matrix is

$$\Sigma_\theta = \Phi_{0\perp}(\Phi'_{0\perp} J_0 \Phi_{0\perp})^{-1} \Phi'_{0\perp} \Sigma_S \Phi_{0\perp} (\Phi'_{0\perp} J_0 \Phi_{0\perp})^{-1} \Phi'_{0\perp},$$

$\Sigma_S = \mathrm{var}\big(\sqrt{n} S_n(\theta_0)\big)$, $\Phi_0 = \Phi(\theta_0)$, and $\Phi_{0\perp}$ is the orthogonal complement of Φ_0 (i.e., $\Phi'_{0\perp}\Phi_0 = 0$ and $[\Phi_{0\perp}, \Phi_0]$ is invertible). The above results also hold when u_i in $s_i(\theta_0)$ is replaced by $\hat{u}_i$. □

The proposition is proven in the Appendix. Note that the asymptotic covariance Σ_θ is of reduced rank. Its rank is at most that of $\Phi_{0\perp}$, $rK - r(r+1)/2 + K_\beta$. When J_0 is invertible (which requires $\mathbb{E}(f_{0k,i}) \neq 0$ for all k), it can alternatively be expressed as

$$\Sigma_\theta = \big[J_0^{-1} - J_0^{-1}\Phi_0(\Phi'_0 J_0^{-1} \Phi_0)^{-1} \Phi'_0 J_0^{-1}\big] \Sigma_S \big[J_0^{-1} - J_0^{-1}\Phi_0(\Phi'_0 J_0^{-1} \Phi_0)^{-1} \Phi'_0 J_0^{-1}\big].$$

This formula can be used when J_0 is singular, by replacing J_0 with $J_{0+} = J_0 + \Phi_0\Phi'_0$, as suggested by Silvey (1959). It can be shown that the above formula with J_{0+} is also equivalent to the formula given in Proposition 2 (see Lemma 3 in the Appendix and note that $\Phi'_{0\perp} J_{0+} \Phi_{0\perp} = \Phi'_{0\perp} J_0 \Phi_{0\perp}$). The matrix Φ_0 is naturally estimated by $\Phi(\hat{\theta})$, i.e.,

$$\Phi(\hat{\theta}) = \begin{bmatrix} (I_r \otimes \hat{\Omega}\hat{A}'_1)\mathcal{D}_r^{+\prime} \\ 0 \end{bmatrix}.$$

Furthermore, $\Phi_{0\perp}$ can be estimated explicitly as

$$\Phi_\perp(\hat{\theta}) = \begin{bmatrix} (I_r \otimes \hat{A}'_1)\mathcal{D}_{r\perp} & (I_r \otimes \hat{A}'_2) & 0 \\ 0 & 0 & I_{K_\beta} \end{bmatrix},$$

where $\mathcal{D}_{r\perp}$ is the orthogonal complement of $\mathcal{D}_r$. When $r = 1$, $\Phi(\hat{\theta})$ and $\Phi_\perp(\hat{\theta})$ simplify to

$$\Phi(\hat{\theta}) = \begin{bmatrix} \hat{\Omega}\hat{a}_1 \\ 0 \end{bmatrix} \quad \text{and} \quad \Phi_\perp(\hat{\theta}) = \begin{bmatrix} \hat{A}_2' & 0 \\ 0 & I_{K_\beta} \end{bmatrix}.$$

When $r = K$ (with $A_1 = A$), $\Phi(\hat{\theta})$ and $\Phi_\perp(\hat{\theta})$ become

$$\Phi(\hat{\theta}) = \begin{bmatrix} (I_K \otimes \hat{\Omega}\hat{A}')\mathcal{D}_K^{+\prime} \\ 0 \end{bmatrix} \quad \text{and} \quad \Phi_\perp(\hat{\theta}) = \begin{bmatrix} (I_K \otimes \hat{A}')\mathcal{D}_{K\perp} & 0 \\ 0 & I_{K_\beta} \end{bmatrix}.$$

It can be shown that $\Phi_{0\perp}' J_0 \Phi_{0\perp}$ is block diagonal and positive definite when $J_{0,22}$ in (10) is invertible (see Lemma 1 in the Appendix). The structure of $\mathcal{D}_{r\perp}$ is also given in the Appendix.

Under the assumptions of Proposition 2, the asymptotic covariance matrix of $S_n(\theta_0)$, Σ_S, is consistently estimated by the outer product form $\hat{\Sigma}_S = n^{-1}\sum_{i=1}^n s_i(\hat{\theta})s_i(\hat{\theta})'$. Hence, the asymptotic covariance of $\hat{\theta}$ is consistently estimated by

$$\hat{\Sigma}_\theta = \hat{\Phi}_\perp(\hat{\Phi}_\perp' \hat{J}\hat{\Phi}_\perp)^{-1}\hat{\Phi}_\perp' \hat{\Sigma}_S \hat{\Phi}_\perp(\hat{\Phi}_\perp' \hat{J}\hat{\Phi}_\perp)^{-1}\hat{\Phi}_\perp', \tag{11}$$

where $\hat{\Phi}_\perp = \Phi_\perp(\hat{\theta})$, and

$$\hat{J} = \frac{1}{n}\sum_{i=1}^n \begin{bmatrix} (\Lambda_i^{-1} - I_r) \otimes u_i u_i' & -(\Lambda_i^{-1} \otimes u_i u_i')\mathrm{diag}(a_1 f_{1,i}', \ldots, a_r f_{r,i}') \\ -(\Lambda_i^{-1} \otimes u_i u_i')\mathrm{diag}(f_{1,i}a_1', \ldots, f_{r,i}a_r') & \frac{1}{2}\mathrm{diag}(f_{1,i}f_{1,i}', \ldots, f_{r,i}f_{r,i}') \end{bmatrix}$$

evaluated at $\hat{\theta}$. This is clearly a "sandwich" form that takes into account the singularity of Σ_S.

For a heteroskedasticity rank $r < K - 1$, the following proposition shows that $\hat{A}_2$, as defined in the estimation procedure at the end of Section 3.1, converges to a rotation of A_{02} at the rate of $n^{-1/2}$.

Proposition 3. *Suppose that the assumptions of Proposition 2 hold. Assume further that A5 holds for* $v_i \equiv \mathrm{vech}(u_i u_i' - \Omega_0)$. *Then,* $\hat{A}_2$ *is asymptotically normal and its distribution is determined by*

$$\sqrt{n}(\hat{A}_2'\hat{d}_2^{-1} - A_{02}') = -A_{01}'(\hat{A}_1\hat{\Omega}A_{01}')^{-1}\sqrt{n}\Big[(\hat{A}_1 - A_{01})\hat{\Omega} + A_{01}(\hat{\Omega} - \Omega_0)\Big]A_{02}',$$

where $\hat{A}_1' = [\hat{a}_1, \ldots, \hat{a}_r]$ *and* $\hat{d}_2 = (A_{02}\Omega_0\hat{A}_2')$. *This result also holds when* u_i *is replaced by* $\hat{u}_i$ *in computing* $\hat{\Omega}$. □

This proposition is also proven in the Appendix. We note that $\hat{d}_2$ does not necessarily converge to an identity matrix because A'_{02} is not identified and $\hat{A}'_2$ can only estimate the space spanned by the columns of A'_{02}. Because $\hat{d}_2$ is unknown in practice, this result cannot be used for point inference about A_{02}. However, it can be used to make inference about the space spanned by the columns of A'_{02}. In particular, Proposition 3 implies that $\hat{A}_2 u_i = \hat{d}'_2 A_{02} u_i + O_p(n^{-1/2}) A_{01} u_i = \hat{d}'_2 \varepsilon_i^{(r+1:K)} + O_p(n^{-1/2}) \varepsilon_i^{(1:r)}$, which will be used in the residual-based heteroskedasticity rank test discussed in Section 4.

3.4 Inference about Coefficients on x_i

As $\hat{A}_1$ and $\hat{D}$ are consistent estimators of A_{01} and $D_0 = A_0^{-1} C_0$ respectively, the first r rows of C_0, denoted as C_{01}, can be consistently estimated by $\hat{C}_1 = \hat{A}_1 \hat{D}$. Given the joint asymptotic distribution of $(\hat{A}_1, \hat{D})$, the delta method delivers the asymptotic distribution of $\hat{C}_1$.

Proposition 4. *Suppose that the assumptions of Proposition 2 hold. Then, the asymptotic distribution of $\hat{C}_1$ is given by*

$$\sqrt{n}\,\mathrm{vec}(\hat{C}'_1 - C'_{01}) \xrightarrow{d} N(0, \Upsilon \Sigma \Upsilon'),$$

where $\Upsilon = [(I_r \otimes D'), 0, (A_1 \otimes I_{K_x})]$ is the Jacobian of $C'_1 = D' A'_1$ with respect to $[\theta', \mathrm{vec}(D')']'$ with 0 being a $rK_x \times K_\beta$ zero matrix corresponding to the $[\beta'_1, \ldots, \beta'_r]'$ part of θ, and Σ is the joint asymptotic covariance of $\sqrt{n}[(\hat{\theta} - \theta_0)', \mathrm{vec}(\hat{D}' - D'_0)']'$. □

The joint asymptotic covariance of $\sqrt{n}[(\hat{\theta} - \theta_0)', \mathrm{vec}(\hat{D}' - D'_0)']'$ may be estimated by

$$\hat{\Sigma} = \begin{bmatrix} \hat{\Sigma}_\theta & \hat{\Sigma}_{\theta D} \\ \hat{\Sigma}'_{\theta D} & \hat{\Sigma}_D \end{bmatrix}, \tag{12}$$

where $\hat{\Sigma}_\theta$ is given by (11),

$$\hat{\Sigma}_D = \left[I_K \otimes \Big(\frac{1}{n} \sum_{i=1}^{n} x_i x'_i \Big)^{-1} \right] \left[\frac{1}{n} \sum_{i=1}^{n} (u_i u'_i \otimes x_i x'_i) \right] \left[I_K \otimes \Big(\frac{1}{n} \sum_{i=1}^{n} x_i x'_i \Big)^{-1} \right],$$

$$\hat{\Sigma}_{\theta D} = \hat{\Phi}_\perp (\hat{\Phi}'_\perp \hat{J} \hat{\Phi}_\perp)^{-1} \hat{\Phi}'_\perp \left[\frac{1}{n} \sum_{i=1}^{n} s_i(\hat{\theta}) \mathrm{vec}(x_i u'_i) \right] \left[I_K \otimes \Big(\frac{1}{n} \sum_{i=1}^{n} x_i x'_i \Big)^{-1} \right].$$

For $K = 1$, $\hat{\Sigma}_D$ is the conventional heteroskedasticity-robust variance estimate in a scalar linear regression model. Clearly, the asymptotic covariance of $\sqrt{n}\,\mathrm{vec}(\hat{C}'_1 - C'_{01})$ is estimated by $\hat{\Sigma}_C = \hat{\Upsilon} \hat{\Sigma} \hat{\Upsilon}'$ with $\hat{\Upsilon} = [(I_r \otimes \hat{D}'), 0, (\hat{A}_1 \otimes I_{K_x})]$. Inference on C_{01} proceeds in a standard manner.

4 Testing for the Heteroskedasticity Rank

For the model defined by (1) and (2), a key parameter is the heteorskedasticity rank r, which is the number of linearly independent conditional variances in the structural error ε_i. It determines the number of rows in A that can be consistently estimated. Given the importance of the heteroskedasticity rank in our model, we now consider testing hypotheses about r. Specifically, we derive tests for the pair of hypotheses $\mathbb{H}_0 : r = r_0$ versus $\mathbb{H}_1 : r > r_0$. Under $\mathbb{H}_0$, the parameter β_k is constrained at a particular point β_{H_0}, typically zero, such that $\sigma^2_{k,i} = 1$. Hence $\mathbb{H}_0$ is equivalent to $\beta_k = \beta_{H_0}$ for all $k \in \{r_0 + 1, \ldots, K\}$. Under $\mathbb{H}_0$, the criterion functions

$$\ell_{k,n}(\beta_k, a_k) = -\frac{1}{n}\sum_{i=1}^{n}\Big[\ln(\sigma^2_{k,i}) + a_k' u_i u_i' a_k(\sigma^{-2}_{k,i} - 1)\Big], \qquad k = r_0 + 1, \ldots, K,$$

are equal to zero and unrelated to $A_2' = [a_{r_0+1}, \ldots, a_K]$. Thus, A_2 is unidentified under $\mathbb{H}_0$, although $\hat{A}_1 \Omega A_2' = 0$ and $A_2 \Omega A_2' = I_{K-r_0}$ must hold for $\hat{A}_1' = [\hat{a}_1, \ldots, \hat{a}_{r_0}]$. This falls into the class of testing problems considered by Davies (1977, 1987), where the nuisance parameter A_2 is present only under $\mathbb{H}_1$. Davies' problem in general settings is considered by Hansen (1996) and Andrews and Ploberger (1994, 1995). We use these ideas to develop likelihood based tests in the following. As these tests may be difficult to implement in practice, we also consider simpler residual-based tests which are easy to conduct in practice.

4.1 Likelihood-Based Tests

In what follows, let $k = r_0 + 1$ and denote the parameter space of a_k as $\Pi_k = \{a : a = Q_k \rho,\ \rho'\rho = 1\}$, where $Q_k = \hat{A}_2'$. Under $\mathbb{H}_0$, Lemma 4 in the Appendix shows that Π_k converges in probability to $\Pi_{0k} = \{a : a = A_{02}'\rho,\ \rho'\rho = 1\}$ in that $A_{01}\Omega_0 Q_k \xrightarrow{p} 0$ and $A_{02}\Omega_0 Q_k \xrightarrow{p} \delta$, where δ is an orthogonal matrix. Here, $\ell_{k,n}$ implicitly depends on $\hat{\theta} = [\text{vec}(\hat{A}_1')', \hat{\beta}_1', \ldots, \hat{\beta}_{r_0}']'$ as Π_k depends on $\hat{\theta}$.

The likelihood ratio statistic is the sup-LR test discussed by Hansen (1996) and Andrews and Ploberger (1995). In our setting, it is simply the maximum of $n\ell_{k,n}(\beta_k, a_k)$,

$$supLR_n = \max_{\beta_k, a \in \Pi_k} n\ell_{k,n}(\beta_k, a) = \max_{\beta_k, a \in \Pi_k}\left\{-\sum_{i=1}^{n}\Big[\ln(\sigma^2_{k,i}) + a' u_i u_i' a(\sigma^{-2}_{k,i} - 1)\Big]\right\},$$

which is readily obtained from the procedure discussed in Section 3. We also define

$supLM_n = \sup_{a\in\Pi_k} LM_n(a)$ with $LM_n(a) = n\mathcal{S}(\beta_{H_0}, a)'[\mathcal{V}_n(\beta_{H_0}, a)]^{-1}\mathcal{S}(\beta_{H_0}, a)$,

$$\mathcal{S}_n(\beta_k, a) = -\frac{\partial \ell_{k,n}}{\partial \beta_k} = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \sigma_{k,i}^{-2} a' u_i u_i' a\right) f_{k,i},$$

and

$$\mathcal{V}_n(\beta_k, a) = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \sigma_{k,i}^{-2} a' u_i u_i' a\right)^2 f_{k,i} f_{k,i}',$$

where $a \in \Pi_k$. Under $\mathbb{H}_1$, we find that, at θ_0 (the true parameter point of the first r_0 equations in (1)), $\mathbb{E}V_n(\beta_{H_0}, a)$ is positive definite and

$$\mathbb{E}S_n(\beta_{H_0}, a) = \mathbb{E}\big(1 - \rho' A_{02} u_i u_i' A_{02}' \rho\big) f_{0k,i} = \mathbb{E}\big(1 - \rho' \varepsilon_i^{(r_0+1:K)} \varepsilon_i^{(r_0+1:K)\prime} \rho\big) f_{0k,i} \neq 0$$

for any $a \in \Pi_{0k}$, as $\varepsilon_i^{(r_0+1:K)}$ is heteroskedastic. It follows that $supLM_n$ diverges to infinity in probability as $n \to \infty$. Under $\mathbb{H}_1$, the results of Section 3 imply that $supLR_n$ also diverges to infinity as $n \to \infty$. Thus, the main goal of this subsection is to find the asymptotic null distributions of $supLR_n$ and $supLM_n$.

For given $a_k = a$, $\hat{\beta}_k(a) = \arg\max_{\beta_k} \ell_{k,n}(\beta_k, a)$ is a function of a, and so are the likelihood ratio statistic $LR_n(a) = \max_{\beta_k} n\ell_{k,n}(\beta_k, a)$ and the LM statistic $LM_n(a)$. Asymptotically, $LR_n(a)$ and $LM_n(a)$ converge weakly to stochastic processes indexed by a. Then the asymptotic null distributions of $supLR_n = \max_{a\in\Pi_k} LR_n(a)$ and $supLM_n = \max_{a\in\Pi_k} LM_n(a)$ are the distributions of the suprema of these stochastic processes. The general distribution theory under high-level assumptions is given by Andrews and Ploberger (1994, 1995) for correctly specified likelihoods, and by Hansen (1996) for possibly misspecified likelihoods. Andrews and Ploberger (1995) show that $supLR_n$ is an admissible test when the likelihood is correctly specified.

Andrews and Ploberger (1994) also consider the following version of the test statistic,

$$expLR_n \quad = \quad (1+c)^{-\frac{1}{2}K_{\beta_k}} \int_{\Pi_k} \exp\Big(\frac{c}{2(1+c)} LR_n(a)\Big) d\mathcal{W}(a),$$

where K_{β_k} is the dimension of β_k, $\mathcal{W}$ is a weight function over Π_k and $c > 0$ is a scalar that controls whether the power is directed toward remote (large c) or local (small c) alternatives. Andrews and Ploberger (1994) show optimality properties of the test based on $expLR_n$. When $c \to \infty$, $expLR_n$ is equivalent to $\ln \int_{\Pi_k} \exp\big(\frac{1}{2}LR_n(a)\big) d\mathcal{W}(a)$. The $expLR_n$ test is closely related to the Bayes factor for Davies' problem (see Yang (2014)). For a correctly specified likelihood, the optimality of $expLR_n$ carries over to the similar version of the LM (or Wald) statistic.

To derive the asymptotic null distribution of $supLR_n$ and $supLM_n$ with primitive conditions in our context, we apply WLLN to $(\mathcal{J}_n, \mathcal{V}_n)$ and CLT to $\mathcal{S}_n$, where $\mathcal{J}_n$ is defined in the Appendix. The required conditions are listed below, where $\|\cdot\|$ stands for the Euclidean norm and $\mathcal{N}_{\beta_{H_0}}$ is a compact neighborhood of β_{H_0} at θ_0.

Assumption B

B1 A4 holds for $k = r_0 + 1$.

B2 There exist functions $g_1(z_i)$ and $g_2(z_i)$ such that $\|f_{k,i} f'_{k,i}\| \leq g_1(z_i)$ and $\|\partial f_{k,i}/\partial \beta'_k\| \leq g_2(z_i)$ for all $\beta_k \in \mathcal{N}_{\beta_{H_0}}$. $\mathbb{E}\big[u'_i u_i \exp(g(z_i)) g_2(z_i)\big]$ and $\mathbb{E}\big[(u'_i u_i)^2 \exp(2g(z_i)) g_1(z_i)\big]$ are finite, where $g(z_i)$ is as defined in A4. Furthermore, $\mathbb{E}(f_{0k,i} f'_{0k,i})$ is invertible.

B3 A5 holds for $v_i \equiv \big(1 - a' u_i u'_i a\big) f_{0k,i}$ for any $a \in \Pi_{0k}$, where $f_{0k,i}$ is $f_{k,i}$ evaluated at β_{H_0}.

B4 A5 holds for $\text{vec}(u_i x'_i) f^{(j)}_{k,i} \sigma^{-2}_{k,i}$, $\text{vec}(u_i x'_i) f^{(j)}_{k,i} f^{(l)}_{k,i} \sigma^{-2}_{k,i}$ and $\text{vec}(u_i x'_i)(\partial f^{(j,l)}_{k,i}/\partial \beta'_k) \sigma^{-2}_{k,i}$ for all $\beta_k \in \mathcal{N}_{\beta_{H_0}}$, where $f^{(j)}_{k,i}$ is the j^{th} element of $f_{k,i}$ and $\partial f^{(j,l)}_{k,i}/\partial \beta'_k$ is the $(j,l)^{\text{th}}$ element of $\partial f_{k,i}/\partial \beta'_k$.

Here, B1 and B2 are needed to apply the uniform WLLN to $\ell_{k,n}$ and $(\mathcal{J}_n, \mathcal{V}_n)$ respectively. B3 allows us to apply a CLT to $\mathcal{S}_n$. The effect of replacing u_i by $\hat{u}_i$ in our analysis becomes negligible under B4. The asymptotic null distributions of $supLR_n$ and $supLM_n$ are given in the following proposition which is proven in the Appendix.

Proposition 5. *Suppose that the assumptions of Proposition 3 hold. Assume further that B1, B2, and B3 hold. Then, under* $\mathbb{H}_0$*,*

(a) $\sqrt{n}\mathcal{S}_n(\beta_{H_0}, a) \Rightarrow \mathcal{S}(a)$ *on* $a \in \Pi_{0k}$*, where* $\mathcal{S}(a)$ *is a zero-mean Gaussian process with covariance function* $\mathcal{K}(a,b) = n\mathbb{E}\big[\mathcal{S}_n(\beta_{H_0}, a)\mathcal{S}_n(\beta_{H_0}, b)'\big]$*;*

(b) $LR_n(a) \Rightarrow \frac{1}{2}\mathcal{S}(a)'[\mathbb{E}(f_{0k,i} f'_{0k,i})]^{-1}\mathcal{S}(a)$ *on* $a \in \Pi_{0k}$*;*

(c) $supLR_n \xrightarrow{d} \sup_{a \in \Pi_{0k}} \frac{1}{2}\mathcal{S}(a)'[\mathbb{E}(f_{0k,i} f'_{0k,i})]^{-1}\mathcal{S}(a)$*;*

(d) $LM_n(a) \Rightarrow \mathcal{S}(a)'[\mathcal{K}(a,a)]^{-1}\mathcal{S}(a)$ *on* $a \in \Pi_{0k}$*;*

(e) $supLM_n \xrightarrow{d} \sup_{a \in \Pi_{0k}} \mathcal{S}(a)'[\mathcal{K}(a,a)]^{-1}\mathcal{S}(a)$*.*

Moreover, under B4, the above results hold when u_i *is replaced by* $\hat{u}_i$*.* □

The asymptotic null distributions of $expLR_n$ and other test statistics suggested by Andrews and Ploberger (1994) can be readily obtained under the assumptions of the above proposition. For example,

$$expLR_n \overset{d}{\to} (1+c)^{-\frac{1}{2}K_{\beta_k}} \int_{\Pi_{0k}} \exp\Big(\frac{c}{4(1+c)}\mathcal{S}(a)'[\mathbb{E}(f_{0k,i}f_{0k,i}')]^{-1}\mathcal{S}(a)\Big)d\mathcal{W}(a).$$

For our setting, the covariance of $\mathcal{S}(a)$ can be expressed as

$$\mathcal{K}(a,b) = \mathbb{E}\big[(1-a'u_iu_i'a)f_{0k,i}f_{0k,i}'(1-b'u_iu_i'b)\big], \qquad a,b \in \Pi_{0k}.$$

In particular, the variance of $\mathcal{S}(a)$ can be simplified as

$$\mathcal{K}(a,a) = \mathbb{E}\Big[\Big(1+(\rho'\varepsilon_i^{(r_0+1:K)})^4 - 2(\rho'\varepsilon_i^{(r_0+1:K)})^2\Big)f_{0k,i}f_{0k,i}'\Big], \qquad \rho'\rho = 1.$$

When the likelihood is correctly specified with $\eta_i = H_i^{-1/2}\varepsilon_i \sim N(0, I_K)$, the variance is $\mathcal{K}(a,a) = 2\mathbb{E}(f_{0k,i}f_{0k,i}')$ and the process $\frac{1}{2}\mathcal{S}(a)'[\mathbb{E}(f_{0k,i}f_{0k,i}')]^{-1}\mathcal{S}(a)$ becomes a χ^2 process on Π_{0k}, which is a χ^2 random variable for any given $a \in \Pi_{0k}$. However, for the QML where η_i is non-normal, $\frac{1}{2}\mathcal{S}(a)'[\mathbb{E}(f_{0k,i}f_{0k,i}')]^{-1}\mathcal{S}(a)$ is generally not a χ^2 process. On the other hand, the asymptotic version of $LM_n(a)$, $\mathcal{S}(a)'\mathcal{K}(a,a)^{-1}\mathcal{S}(a)$, is always a χ^2 process on Π_{0k}. The sup-Wald statistic has the same asymptotic null distribution as $supLM_n$ when the sandwich-form covariance matrix is used. The LM test is particularly simple for our purpose as it can be carried out without estimating β_k. The asymptotic null distribution of $supLR_n$ (or $supLM_n$) depends on nuisance parameters (i.e., A_2). Tabulating critical values is not feasible. Hansen (1996) suggests a simulation procedure to compute the asymptotic p-value of $supLM_n$, which is summarized for our setting as follows.

For any $a \in \Pi_k$, let $v_i(a) = \big(1 - a'u_iu_i'a\big)f_{0k,i}$ be the (observable) summand in $\mathcal{S}_n(\beta_{H_0}, a)$. Draw an independent sample $\{\omega_i\}_{i=1}^n$ from $N(0,1)$. Construct the simulated process $\tilde{\mathcal{S}}_n(a) = n^{-1/2}\sum_{i=1}^n v_i(a)\omega_i$, which is a zero-mean Gaussian process with covariance $\tilde{\mathcal{K}}(a,b) = n^{-1}\sum_{i=1}^n v_i(a)v_i(b)'$, conditional on $\mathcal{F}_n$. Find the simulated test statistic $\tilde{T}_n^{(1)} = \max_{a\in\Pi_k}\tilde{\mathcal{S}}_n(a)'[\tilde{\mathcal{K}}(a,a)]^{-1}\tilde{\mathcal{S}}_n(a)$. Repeat this procedure N times to obtain $\{\tilde{T}_n^{(t)}\}_{t=1}^N$. Compute $\tilde{p} = N^{-1}\sum_{t=1}^N \mathbf{1}(\tilde{T}_n^{(t)} > supLM_n)$ as the p-value estimate, where $\mathbf{1}(\cdot)$ is the indicator function.

The procedure is valid because $\tilde{\mathcal{K}}(a,b) \overset{p}{\to} \mathcal{K}(a,b)$ and $\tilde{\mathcal{S}}_n(a) \Rightarrow \mathcal{S}(a)$. The implementation of $supLM_n$ requires a maximization over Π_k for each simulated sample. Clearly, these tests are difficult to conduct in practice. More pragmatic tests are considered in the following.

4.2 Residual-Based Tests

In this subsection we use the notation $\varepsilon_i^{(1)} = \varepsilon_i^{(1:r_0)}$, $\varepsilon_i^{(2)} = \varepsilon_i^{(r_0+1:K)}$, and $\tau = K - r_0$. Under $\mathbb{H}_0$, $A_{02}u_i = \varepsilon_i^{(2)}$ is the homoskedastic part of the structural error ε_i and $\mathbb{E}(A_{02}u_iu_i'A_{02}'|W_i) = I_\tau$ does not depend on W_i. Defining $\xi_i = \text{vech}(A_{02}u_iu_i'A_{02}') = \text{vech}(\varepsilon_i^{(2)}\varepsilon_i^{(2)\prime})$, the parameter α_1 in the "ideal" regression

$$\xi_i = \alpha_0 + \alpha_1 w_i + \zeta_i, \tag{13}$$

is zero, i.e. $\alpha_1 = 0$, under $\mathbb{H}_0$. Here the regressor $w_i \in W_i$ is a vector of exogenous non-constant variables and ζ_i is the error term. The vector w_i typically contains known functions of x_i or z_i. Under $\mathbb{H}_0$, $\alpha_0 = \text{vech}(I_r)$ and $\zeta_i = \text{vech}(\varepsilon_i^{(2)}\varepsilon_i^{(2)\prime} - I_\tau)$. Under $\mathbb{H}_1$, $\mathbb{E}(\xi_i|W_i)$ contains non-trivial conditional variances of $\varepsilon_i^{(2)}$. Hence, the estimator of α_1 converges in probability to zero under $\mathbb{H}_0$, and to a non-zero constant matrix under $\mathbb{H}_1$ when w_i is properly chosen and correlated with the conditional variances. Thus, a Wald test is informative on the hypothesis $\alpha_1 = 0$. Clearly, a rejection of $\alpha_1 = 0$ is a rejection of $\mathbb{H}_0$. However, not rejecting $\alpha_1 = 0$ does not imply that $\mathbb{H}_0$ cannot be rejected, unless $\text{cov}(\xi_i, w_i) \neq 0$ under $\mathbb{H}_1$.

As A_{02} is unknown, we may replace it with $\hat{A}_2$, which is defined in Section 3 (also see Proposition 3). Let $\hat{\xi}_i = \text{vech}(\hat{A}_2u_iu_i'\hat{A}_2')$, where we still use u_i instead of $\hat{u}_i$. The effect of using $\hat{u}_i$ will be assessed later on. We consider the feasible OLS estimator of $\alpha = [\alpha_0, \alpha_1]$ in (13), using $\hat{\xi}_i$,

$$\hat{\alpha} = [\hat{\alpha}_0, \hat{\alpha}_1] \;=\; \Big(\sum_{i=1}^n \hat{\xi}_i Z_i'\Big)\Big(\sum_{i=1}^n Z_iZ_i'\Big)^{-1},$$

where $Z_i' = [1, w_i']$. The asymptotic covariance of $\text{vec}(\hat{\alpha})$ is estimated as

$$\hat{V}_\alpha = \Big(\frac{1}{n}\sum_{i=1}^n Z_iZ_i'\Big)^{-1} \otimes \Big(\frac{1}{n}\sum_{i=1}^n \hat{\zeta}_i\hat{\zeta}_i'\Big),$$

where $\hat{\zeta}_i = \hat{\xi}_i - \hat{\alpha}_0 - \hat{\alpha}_1 w_i$. Let K_w be the dimension of w_i and $h = [0, I_{K_w}]'$ such that $\hat{\alpha}h = \hat{\alpha}_1$. Then the Wald statistic for testing $\alpha_1 = 0$ can be expressed as

$$\text{Wald}_{1,n} = n\,\text{vec}(\hat{\alpha}h)' \left[h'\Big(\frac{1}{n}\sum_{i=1}^n Z_iZ_i'\Big)^{-1} h \otimes \Big(\frac{1}{n}\sum_{i=1}^n \hat{\zeta}_i\hat{\zeta}_i'\Big)\right]^{-1} \text{vec}(\hat{\alpha}h).$$

An alternative regression for testing $\mathbb{H}_0$ is obtained by using the sum of the squared errors $\varepsilon_i^{(2)\prime}\varepsilon_i^{(2)} = u_i'A_2'A_2u_i = q'\xi_i$ as the dependent variable. Here $q = [e_\tau^{1\prime}, \ldots, e_2^{1\prime}, 1]'$, where e_k^1 is the first column of I_k for $k \in \{\tau, \tau-1, \ldots, 2\}$. Using the univariate regression

$$q'\xi_i = q'\alpha[1, w_i']' + q'\zeta_i,$$

the coefficients on w_i are zero under $\mathbb{H}_0$ and nonzero under $\mathbb{H}_1$, provided $\text{cov}(q'\xi_i, w_i) \neq 0$. We again substitute $\hat{\xi}_i$ for ξ_i for actually performing the regression and computing the Wald statistic for $q'\alpha_1 = 0$. The Wald statistic in this scalar regression can alternatively be written as

$$\text{Wald}_{2,n} = n\ (q'\hat{\alpha}h) \left[h'\Big(\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i'\Big)^{-1} h \otimes q'\Big(\frac{1}{n}\sum_{i=1}^{n} \hat{\zeta}_i\hat{\zeta}_i'\Big) q \right]^{-1} (q'\hat{\alpha}h)'.$$

This expression is useful for finding the asymptotic properties of the test statistic $\text{Wald}_{2,n}$ via those of $\hat{\alpha}$. The following proposition provides the asymptotic properties of $\hat{\alpha}$ and the Wald tests.

Proposition 6. *Suppose that the assumptions of Proposition 3 hold. Assume further that* $V_Z = \mathbb{E}(Z_i Z_i')$ *is invertible, A5 holds for* $v_i \equiv \text{vec}\big(\text{vec}(\varepsilon_i^{(1)}\varepsilon_i^{(2)\prime})Z_i'\big)$ *and* $v_i \equiv \text{vec}\big(\zeta_i Z_i'\big)$*, where* $\zeta_i = \text{vech}(\varepsilon_i^{(2)}\varepsilon_i^{(2)\prime} - I_\tau)$*. Then the following results hold.*

(a) Under $\mathbb{H}_0$, $\sqrt{n}\,\text{vec}\Big([\hat{\alpha}_0, \hat{\alpha}_1] - M_n[\alpha_0, \alpha_1]\Big) \xrightarrow{d} N(0, V_\alpha)$*, where* $V_\alpha = V_Z^{-1} \otimes M_0 V_\zeta M_0'$, $V_\zeta = \text{var}(\zeta_i)$, $M_n = \mathcal{D}_\tau^{+}(\hat{d}_2' \otimes \hat{d}_2')\mathcal{D}_\tau$, $M_0 = \mathcal{D}_\tau^{+}(\delta' \otimes \delta')\mathcal{D}_\tau$, $\hat{d}_2$ *is defined in Proposition 3,* δ *is an orthogonal matrix,* $\alpha_1 = 0$ *and* $\alpha_0 = \mathbb{E}(\xi_i)$.

(b) Under $\mathbb{H}_0$, $\hat{V}_\alpha \xrightarrow{p} V_\alpha$.

(c) Under $\mathbb{H}_0$, $\text{Wald}_{1,n} \xrightarrow{d} \chi^2\big(\tfrac{1}{2}\tau(\tau+1)K_w\big)$ *and* $\text{Wald}_{2,n} \xrightarrow{d} \chi^2(K_w)$.

(d) Under $\mathbb{H}_1$, $[\hat{\alpha}_0, \hat{\alpha}_1] \xrightarrow{p} M_0[\alpha_0, \alpha_1]$*, where* $\alpha_1 = C_{\xi,w}V_w^{-1}$, $\alpha_0 = \mathbb{E}(\xi_i) - \alpha_1\mathbb{E}(w_i)$, $V_w = \text{var}(w_i)$ *and* $C_{\xi,w} = \text{cov}(\xi_i, w_i')$.

(e) Under $\mathbb{H}_1$, $\hat{V}_\alpha \xrightarrow{p} V_Z^{-1} \otimes M_0(V_\xi - C_{\xi,w}V_w^{-1}C_{\xi,w}')M_0'$*, where* $V_\xi = \text{var}(\xi_i)$.

(f) Under $\mathbb{H}_1$, $n^{-1}\text{Wald}_{1,n} \xrightarrow{p} c_1$ *and* $n^{-1}\text{Wald}_{2,n} \xrightarrow{p} c_2$*, where* (c_1, c_2) *are constants,* $c_1 > 0$ *if* $C_{\xi,w} \neq 0$*, and* $c_2 > 0$ *if* $q'C_{\xi,w} \neq 0$.

Moreover, if $\mathbb{E}\big(\text{vec}(x_i x_i)Z_i'\big)$ *is finite and A5 holds for* $v_i \equiv \text{vec}\big(\text{vec}(u_i x_i)Z_i'\big)$*, then the above results hold when* u_i *is replaced by* $\hat{u}_i$. □

Because $\hat{A}_2$ can only be used to estimate the space spanned by the rows of A_{02}, using $\hat{\xi}_i = \text{vech}(\hat{A}_2 u_i u_i' \hat{A}_2')$ in (13) is markedly different from using the ideal $\xi_i = \text{vech}(A_{02}u_i u_i' A_{02}')$. This difference is reflected in the presence of the matrices M_n and M_0 in Proposition 6 (a) and (d), respectively. Fortunately, this difference does not hinder testing the restriction $\alpha_1 = 0$ because $\hat{\alpha}_1 \xrightarrow{p} 0$ under $\mathbb{H}_0$ and converges to a constant, which is non-zero if

$\text{cov}(\xi_i, w_i') \neq 0$, under $\mathbb{H}_1$. The residual-based Wald tests are pragmatic in the sense that the test statistics are easy to compute and they have standard asymptotic null distributions (χ^2), as indicated in (c). Hence, these tests are easy to implement as they do not require simulation to compute critical values or p-values, whereas simulation is needed for $supLM_n$. The results in (f) imply that the residual-based Wald tests are consistent under $\mathbb{H}_1$ as long as w_i is chosen such that $\text{cov}(\xi_i, w_i') \neq 0$ and $q'\text{cov}(\xi_i, w_i') \neq 0$.

We note that Davies' problem does not show up in the residual-based tests. The proposed tests can be carried out in the standard manner despite the fact that A_2 is not point identified under $\mathbb{H}_0$. The reason is that the implication of $\mathbb{H}_0$ in (13) depends on A_2 via the definition of ξ_i. In other words, in the framework of (13), A_2 is not absent under $\mathbb{H}_0$ as ξ_i is defined in terms of the estimable space spanned by the rows of A_2.

5 Monte Carlo Investigation

We carry out simulation experiments to investigate the finite-sample properties of the proposed QML estimator and heteroskedasticity rank tests. The data generating process (DGP) is inspired by the empirical application in Section 6. Our experiments cover sample sizes $n = 50, 100, 200$ and 500, encompassing the sample size of the data set ($n = 114$) in Section 6. The DGP is the model detailed in (1) and (2) with the dimension $K = 3$ and the exponential functional form is employed to specify the conditional variances. Each sample is generated according to the following steps.

1. Draw independent scalar random numbers $\{w_i\}_{i=1}^n$ from the standard normal distribution $N(0, 1)$ and set $x_i = [1, w_i]'$ and $z_i = w_i$ for all i.

2. Draw 3-dimensional independent random vectors $\{\eta_i\}_{i=1}^n$ from the $\chi^2(9)$ distribution, where the elements of $\eta_i = [\eta_{1,i}, \eta_{2,i}, \eta_{3,i}]'$ are independent and normalized to have mean 0 and variance 1.

3. Generate the conditional variances $\sigma_{k,i}^2 = \exp(\beta_k z_i)/\mathbb{E}\exp(\beta_k z_i)$ and the structural error terms $\varepsilon_{k,i} = \sigma_{k,i}\eta_{k,i}$ for $k = 1, 2, 3$. Set $\varepsilon_i = [\varepsilon_{1,i}, \varepsilon_{2,i}, \varepsilon_{3,i}]'$ for $i = 1, \ldots, n$.

4. Endogenous variables are generated from the reduced-form system $y_i = Dx_i + A^{-1}\varepsilon_i$ based on (1), where $y_i = [y_{1,i}, y_{2,i}, y_{3,i}]'$.

Here, we use $\chi^2(9)$ variates as the standardized structural error term η_i to demonstrate that our QML approach works for non-Gaussian data. The unconditional variances of

$(\varepsilon_{1,i}, \varepsilon_{2,i}, \varepsilon_{3,i})$ in the DGP are normalized to be unity as defined in (1) and (2). The normalizing factor here is $\mathbb{E}\exp(\beta_k z_i) = \exp(\frac{1}{2}\beta_k^2)$. In estimation or testing, we use the specification $\sigma^2_{k,i} = \exp(\beta_k z_i)/[\frac{1}{n}\sum_{i=1}^n \exp(\beta_k z_i)]$ to impose the normalization rule $\mathbb{E}(\sigma^2_{k,i}) = 1$.

The parameter matrices for the DGP are

$$A = \begin{bmatrix} 1.604 & 2.542 & 0.252 \\ -0.280 & 0.604 & 0.896 \\ -0.490 & 5.206 & -0.259 \end{bmatrix}, \qquad D = \begin{bmatrix} 0.0 & 0.2 \\ 0.0 & -0.1 \\ 0.0 & -0.2 \end{bmatrix}.$$

In particular, the instantaneous impact matrix A is the point estimate from the empirical example in Section 6. In the reduced-form coefficient matrix D, the first column corresponds to the intercept and the second to the exogenous variable w_i. While the first column of D is set to zero in the DGP, the intercept is always included in estimation and testing. Hence, changing the values in the first column of D does not alter the results reported below. We use two sets of values for the parameters in the conditional variances: $(\beta_1, \beta_2, \beta_3) = (1, 0, 0)$ and $(1, 0.5, 0)$. The first is the case with the heteroskedasticity rank $r = 1$, for which the model is partially identified. The second is the case with $r = 2$, for which the model is fully identified. All of our experiments consist of 1,500 replications, from which point estimates and test statistics are recorded.

To examine the properties of the QML estimator, we estimate the parameters in the conditional variances and the three rows of A from each sample, and compute bias and root mean squared error (RMSE) from 1,500 replications.[2] Since the estimator of a row in A, $\hat{a}_k$, is consistent up to the scale ± 1, we insist on nonnegative first, third and second elements of $\hat{a}_1, \hat{a}_2$ and $\hat{a}_3$, respectively. For example, if the third element of $\hat{a}_2$ is negative, we will record $-\hat{a}_2$ instead of $\hat{a}_2$ itself. In maximizing the criterion function over the space of β_k, the initial value for the maximization routine is fixed at 0.1.

The estimation results are reported in Tables 1 and 2 for partially and fully identified models respectively. In Table 1, as the model is partially identified, only the first row of A, $a_1' = [a_{11}, a_{12}, a_{13}]$, can be consistently estimated. Indeed, the bias and RMSE of the QML estimator for the first row of A are small and generally decrease as the sample size increases. On the other hand, the bias and RMSE for the second and third rows of A, which are not identified, are large and do not decrease as the sample size increases. We

[2]The computations of the simulation experiments is carried out in R (see R-Team (2016)). Maximizations are done using the BFGS method in the optimization function `optim` of R.

also note that the bias and RMSE are dependent on the magnitude of the true parameter value. For instance, the bias and RMSE of $\hat{a}_{12}$ are larger than those of $\hat{a}_{13}$. In Table 2, as expected from a fully identified model, we observe that the bias and RMSE for all parameters are small and decrease as the sample size increases. Interestingly, for this DGP the element a_{22} appear to be difficult to estimate, having largest RMSE. It corresponds to a coefficient estimate in the application in Section 6, which is the only statistically insignificant element in A (see Table 5).

Table 1: Estimation Bias and RMSE for Partially Identified Model ($r = 1$)

	True	$n = 50$		$n = 100$		$n = 200$		$n = 500$	
	value	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
a_{11}	1.604	0.065	0.343	0.072	0.252	0.027	0.177	0.013	0.114
a_{12}	2.542	0.093	1.129	0.126	0.669	0.044	0.416	0.020	0.262
a_{13}	0.252	0.008	0.180	0.014	0.108	0.006	0.065	0.004	0.039
β_1	1.000	-0.052	0.245	-0.049	0.185	-0.030	0.132	-0.016	0.088
a_{21}	-0.280	0.238	0.675	0.233	0.544	0.215	0.481	0.201	0.451
a_{22}	0.604	-1.231	4.143	-1.384	4.011	-1.303	3.896	-1.245	3.834
a_{23}	0.896	-0.283	0.435	-0.287	0.419	-0.295	0.423	-0.293	0.417
β_2	0.000	0.097	0.337	0.099	0.229	0.068	0.166	0.046	0.107
a_{31}	-0.490	0.162	0.493	0.154	0.381	0.159	0.326	0.150	0.287
a_{32}	5.206	-1.712	2.572	-1.791	2.534	-1.846	2.553	-1.823	2.499
a_{33}	-0.259	0.185	0.735	0.187	0.701	0.182	0.689	0.165	0.670
β_3	0.000	-0.188	0.270	-0.128	0.186	-0.084	0.130	-0.050	0.078

To investigate the finite-sample properties of three proposed tests (Wald_1, Wald_2, supLM), we employ the setup that the first equation in the model is always heteroskedastic. In this scenario, we test $\mathbb{H}_0 : r = 1$ against $\mathbb{H}_1 : r > 1$. Samples from the DGP with $(\beta_1, \beta_2, \beta_3) = (1, 0, 0)$ and $(1, 0.5, 0)$ respectively are used to examine the size and power properties of the tests. A single exogenous variable, w_i, is included in (13) to compute Wald_1 and Wald_2. For the supLM statistic, 100 bootstraps are used to compute its p-value. In maximizing the $LM(a)$ statistic, the initial value for the maximization routine is fixed at 0.1. The rejection rates of the three tests are reported in Table 3, where the first panel ($r = 1$) corresponds to $(\beta_1, \beta_2, \beta_3) = (1, 0, 0)$ and the second and third panels ($r = 2$) correspond to $(\beta_1, \beta_2, \beta_3) = (1, 0.5, 0)$. The first and second panels consist of

Table 2: Estimation Bias and RMSE for Fully Identified Model ($r = 2$)

	True	$n = 50$		$n = 100$		$n = 200$		$n = 500$	
	value	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
a_{11}	1.604	-0.101	0.440	-0.023	0.300	-0.007	0.195	0.004	0.114
a_{12}	2.542	-0.198	1.203	-0.037	0.757	-0.009	0.440	0.006	0.266
a_{13}	0.252	-0.034	0.369	-0.008	0.263	0.002	0.156	0.003	0.083
β_1	1.000	-0.022	0.223	-0.037	0.175	-0.025	0.129	-0.015	0.087
a_{21}	-0.280	0.157	0.885	0.053	0.638	0.002	0.386	-0.003	0.208
a_{22}	0.604	-0.117	2.582	-0.006	1.733	-0.007	1.052	-0.006	0.598
a_{23}	0.896	-0.109	0.310	-0.052	0.207	-0.013	0.115	-0.003	0.058
β_2	0.500	-0.049	0.226	-0.022	0.162	-0.012	0.121	-0.001	0.080
a_{31}	-0.490	0.033	0.482	0.009	0.294	0.002	0.170	-0.003	0.102
a_{32}	5.206	-0.271	1.469	-0.084	0.904	-0.017	0.507	0.011	0.266
a_{33}	-0.259	0.025	0.435	0.001	0.280	0.003	0.175	-0.002	0.094
β_3	0.000	-0.121	0.267	-0.047	0.178	-0.022	0.132	-0.010	0.081

rejection rates based on χ^2 critical values for Wald_1 and Wald_2, and the bootstrap p-value for supLM. The third panel (Power*) contains the size-corrected rejection rates, where the 5% and 10% critical values are the empirical critical values obtained from the simulation in the first panel (under $\mathbb{H}_0$). These entries would be true powers based on the correct (rather than estimated) critical values. They are useful for gauging power distortions caused by size distortions of the tests.

In the first panel of Table 3, the sizes of the three tests are reasonably precise even with the small sample size $n = 50$, indicating that the asymptotic null distributions are good approximations to the finite sample null distributions in this setup. In the second panel of Table 3, we observe that the power of the tests increases toward one as the sample size increases. In terms of power, Wald_1 is ranked best, supLM the second, and Wald_2 the third. An exception is that Wald_2 outperforms supLM at the 5% level when $n = 50$. In the third panel of Table 3, the size-corrected powers do not deviate much from the powers reported in the second panel. This is an indication that the size distortions shown in the first panel do not lead to large power distortions in our simulation experiments.

In summary, the simulation experiments demonstrate that our QML estimator performs effectively in finite samples. They also show that the asymptotic null distributions of

Table 3: Rejection Rates for Testing $\mathbb{H}_0 : r = 1$

		$n = 50$		$n = 100$		$n = 200$		$n = 500$	
Nominal size		5%	10%	5%	10%	5%	10%	5%	10%
$r = 1$	Wald_1	0.077	0.142	0.067	0.119	0.059	0.117	0.054	0.107
(Size)	Wald_2	0.058	0.106	0.051	0.105	0.054	0.102	0.057	0.108
	supLM	0.048	0.105	0.049	0.104	0.035	0.091	0.048	0.095
$r = 2$	Wald_1	0.263	0.367	0.513	0.632	0.841	0.911	0.999	0.999
(Power)	Wald_2	0.175	0.252	0.377	0.497	0.682	0.795	0.977	0.987
	supLM	0.138	0.261	0.410	0.593	0.759	0.885	0.979	0.995
$r = 2$	Wald_1	0.206	0.291	0.462	0.593	0.838	0.898	0.999	0.999
(Power*)	Wald_2	0.164	0.245	0.375	0.492	0.663	0.792	0.973	0.987
	supLM	0.147	0.269	0.423	0.595	0.808	0.891	0.985	0.993

the tests are good approximations to the finite-sample null distributions. Finally, Wald_1, in addition to being easy to implement, has superior power in this setup.

6 Empirical Illustration

We apply our method to re-examine the empirical evidence on openness, inflation and real income provided in Romer (1993). Romer argues that models, in which the absence of precommitment in monetary policy causes excessive inflation, lead to the conclusion that more open economies experience lower average inflation rates. Romer (1993) conducts a cross-country analysis and uses single equation models, where inflation is the dependent variable and openness, real per capita income and possibly other variables serve as explanatory variables. He accounts for potential endogeneity of the explanatory variables by employing instrumental variables (IV) estimation in some of his regressions, and provides evidence in support of this hypothesis. Our aim is to re-cast the analysis in the context of a SEM which allows for possible endogeneity between the three variables. Given that we find conditional heteroskedasticity in the errors, we use our approach to circumvent identification problems.

The relationship between country openness, inflation, and real income is of interest for two reasons. First, Romer (1993) proposes the idea of *endogenous openness* whereby not only is inflation a function of openness, but these two variables are jointly determined through protectionist policies. Extending this argument, one may also conjecture that real income is jointly determined with inflation and openness, and hence is itself endogenous – a possibility not explored in Romer (1993). Second, the analysis is conducted by controlling for a number of exogenous variables, such as the country land area and regional dummy variables, which are used to account for geographical variation in the mean equations. We extend the study by suggesting that these exogenous variables may also drive the conditional variance processes. If that is indeed the case then we may cast the analysis in the HSEM framework described in (1) – (2), which will fully account for any endogeneity that may exist between the three variables. We start the analysis by providing a brief description of the data and testing for the number of heteroskedastic structural innovations, i.e. the heteroskedasticity rank.

Our dataset is obtained from Romer (1993) and consists of several key variables for a cross section of 114 countries. It includes the following three (possibly) endogenous variables:

- π_i – **inflation** as computed by the average annual change in the log GDP or GNP deflator (depending on the availability of data) between 1973 and 1991;
- o_i – country **openness** measured by the average share of imports in GDP or GNP (depending on the availability of data) between 1973 and 1991;
- ry_i – real **income** recorded as the 1980 real income per capita in U.S. dollars.

In addition, we also have data on three exogenous variables:

(i) $Land_i$ – country land area measured as the natural logarithm of the total square miles area for each country;

(ii) I_i^{Am} – a geographical indicator variable set to one for countries located in the Americas region and zero otherwise;

(iii) I_i^{Oil} – oil-producing country indicator variable taking the value of one for oil-producing countries and zero otherwise.

Let $y_i = [\pi_i, o_i, ry_i]'$ and $x_i = [1, Land_i, I_i^{Oil}, I_i^{Am}]'$. In the first step we estimate the reduced-form system (3), and apply our tests for the heteroskedasticity rank, r, on the residuals $\hat{u}_i = y_i - \hat{D}x_i$ as discussed in Section 4, where $\hat{D}$ is obtained via OLS. The log conditional variances, $F_k(z_i, \beta_k)$, are specified as a linear function of $z_i = [Land_i, I_i^{Oil}, I_i^{Am}]'$.

Table 4: Testing for Heteroskedsticity Rank (r)

Test		$\mathbb{H}_0 : r = 0$	$\mathbb{H}_0 : r = 1$
Wald_1	Statistic	47.749	31.285
	p-value	0.000	0.000
Wald_2	Statistic	11.984	8.933
	p-value	0.007	0.030
supLM	Statistic	23.753	22.823
	p-value	0.000	0.000

Notes: In the construction of the test statistics for testing $\mathbb{H}_0 : r = 0$ we set $\hat{\varepsilon}_i^{(2)}$ described in Section 4.2 equal to $\hat{u}_i$. Conditional variance equations are specified as exponential functions of the exogenous vector z_i.

The sample means of the conditional variances are normalized to one in our estimation. Test results for the multivariate (Wald_1) and univariate (Wald_2) Wald tests, as well as the supLM test are reported in Table 4.

Considering the first column of the above table we strongly reject the null hypothesis of no heteroskedasticity in the structural system according to all three tests. This leads us to infer that the heteroskedasticity rank is at least one. Similarly, the results presented in the second column of the table provide evidence against the null hypothesis of one heteroskedastic component ($r = 1$) in favour of the heteroskedasticity rank being at least two ($r \geq 2$), at the 5% level. These results suggest that there is heterogeneity in the variances that can be used for identification. Given that a 3-dimensional system is fully identified with $r \geq 2$, we proceed to estimate all rows of the A matrix using the procedure described in Section 3.1. The estimates are presented in Table 5.

As illustrated in Table 5, all but one coefficient are statistically significant, at the 1% level, which confirms our conjecture that the variable ry_t is indeed determined jointly with π_i and o_i. While each equation of (1) can be consistently estimated in the order of the mean log conditional variances as explained in Sections 3.1 and 3.2, its interpretation depends on the underlying economics. Specifically, to compare our estimates to the results of Romer (1993), we need to decide which of the estimated equations corresponds to his inflation equation. Of course, since we have three equations none of which is economically identified so far, there are three possibilities for the inflation equation. A choice should be based on economic arguments.

Table 5: Estimated Rows of A

Estimated Row	Inflation	Openness	Real Income
Row 1	1.604	2.542	0.252
	[0.000]	[0.000]	[0.000]
Row 2	0.280	−0.604	−0.896
	[0.000]	[0.167]	[0.000]
Row 3	0.490	−5.206	0.259
	[0.000]	[0.000]	[0.006]

Notes: p-values of t-statistics are provided in square brackets.

Following Romer's (1993) argument that inflation is negatively related to openness in his inflation equation, we label the first row of $\hat{A}$ in Table 5 as the inflation equation because it is the only row where the coefficients of inflation and openness have the same sign. Thus, if we standardize the coefficient of inflation to one and use that variable as left-hand side variable with all other variables on the right-hand side, we get an equation where openness reduces inflation. Clearly, the first equation is the only one that gives rise to a negative relationship between inflation and openness. Further, in the absence of additional knowledge about the signs of the coefficients on the endogenous variables, we use a normalization such that the coefficient on real income in the second row is one and the coefficient on openness in the third row is one. This normalization is mainly motivated by the statistical significance of the estimated coefficients. As the coefficient on openness in the second row of A is statistically insignificant, it cannot justifiably be normalized to one. Under this normalization, the results are presented in the first panel of Table 6, where p-values are computed using the delta method.

In Table 6 we also report the results of single equation IV estimation of the inflation equation obtained by utilizing $Land_i$ as an instrument for openness as in Romer (1993). Since this is the only instrument available to Romer, he assumes that income is exogenous in his single-equation model. Note that the p-values of the IV estimates are based on heteroskedasticity-adjusted standard errors.

Considering the coefficients on the openness variable in the two estimated inflation equations, and the associated p-values, we conclude that there is strong support for the hypothesis investigated in Romer (1993), namely that higher levels of economic openness lead to lower inflation rates on average. Negative estimates, statistically significant at

Table 6: Estimated Normalized Structural Relationships

Estimation method	Left-hand side variable	Right-hand side variables: Inflation	Openness	Real Income	White Test on Standardized Residuals
System	Inflation	–	−1.585	−0.157	0.226
		–	[0.000]	[0.000]	[0.973]
	Openness	0.094	–	0.050	0.023
		[0.000]	–	[0.022]	[0.999]
	Real Income	0.313	−0.674	–	0.732
		[0.002]	[0.256]	–	[0.866]
IV	Inflation	–	−1.145	−0.090	–
		–	[0.000]	[0.126]	–

Notes: p-values are provided in square brackets. The covariance matrix of $\hat{A}^*$ computed by sequential estimation is obtained by the delta method from the estimated covariance matrix of $\hat{A}$. The last column of the table provides White's heteroskedasticity test applied to standardized structural residuals (see White (1980)).

the 1% level, are obtained with both estimation methods. Regarding the magnitude of the estimates, we see that the normalized coefficient of −1.585 obtained with our system estimation method is larger in absolute value than the IV estimate of −1.145. These estimates are roughly in line with the results provided in Romer (1993) which range from −0.827 to −1.395, depending on whether or not endogeneity of openness is taken into account and which variables are included in the equation.

The question of whether openness is indeed endogenous can be addressed by examining the estimated coefficient of the inflation variable in the openness equation. As seen in the second row of Table 6, the parameter of inflation is statistically significant at the 1% level in the openness equation, which, taken together with corresponding results from the inflation equation, implies that openness and inflation are indeed jointly determined.

Turning to the relationship between inflation and real income we note substantial differences in the estimated income coefficients in the inflation equation, depending on the estimation method used. The estimate from the system estimation method differs distinctly from the estimate obtained by IV. The system approach estimates a negative and statistically significant (at the 1% level) effect of real income on inflation. While the parameter estimated by IV is also negative, it is almost half the magnitude of the coefficient obtained via the system method and has a much higher p-value of 0.126. The IV finding is similar to the evidence presented in Romer (1993), where the coefficient of real income in the inflation equation is not statistically significant. A reason for this discrepancy can be seen in Table 6, where the system estimates clearly suggest endogeneity of income. The IV estimator ignores the possible endogeneity of income, since the only instrument available in the study is used to instrument openness. This highlights the usefulness of our approach in situations where insufficient identifying information from conventional sources is available. Although the two last equations of our system are not necessarily economically identified and, hence, any economic interpretation can be problematic, they do provide evidence of endogeneity of the income variable.

Lastly, we point out that modeling structural conditional variance equations as exponential functions of exogenous variables accounts for heteroskedasticity adequately. This is judged by White's test for heteroskedasticity (see White (1980)) which is applied here to the standardized structural residuals. The test statistics are small and the associated p-values in excess of 0.850, as reported in the last column of Table 6. This provides confidence in our identification and the estimation approach.

7 Conclusions

This paper presents a complete framework for analysing HSEMs that may be partially identified through (conditional) heteroskedasticity. An estimation method is developed that provides consistent and asymptotically normal estimates of the identified parameters. These results are useful because they can be combined with traditional identification restrictions. In other words, identification through heteroskedasticity can complement identification restrictions from economics. Thus, identification through heteroskedasticity can make up for insufficient identifying economic information. If the combined identification restrictions from traditional sources and heteroskedasticity are overidentifying, this feature can be used to test competing economic hypotheses against the data.

Because identification through heteroskedasticity is linked to the heterogeneity in the variances of the structural errors which we measure by the heteroskedasticity rank, we also develop tests for the heteroskedasticity rank. Thus, we effectively develop tests for identification which inform about the identified equations in the model. These tests can be used even in underidentified or partially identified models. Two alternative asymptotic approaches are used to derive such tests. The first approach is based on Gaussian quasi-likelihood methods and uses techniques that account for nuisance parameters that are only present under the alternative hypothesis. Unfortunately, these tests may not be very practical in many situations because they have nonstandard asymptotic distributions under the null hypothesis and may require a substantial computational effort. Therefore we also derive more practical tests based on Wald principles that have standard asymptotic χ^2-distributions under the null hypothesis and are easy to compute. We compare the two types of tests in a Monte Carlo study and find that the quasi-likelihood based supLM test does not have better power than a Wald type test in small samples. Hence, it may not warrant the additional effort in conducting the supLM type tests, in particular, if full or sufficient identification is found already with the Wald type tests.

We illustrate our approach by reconsidering the question whether openness of an economy is linked to inflation. This issue has been discussed in the literature without fully accounting for possible endogeneity problems related to the variables. Using our approach we can account for the possible endogeneity of the variables. We find support for the theory-based view that openness is negatively related to inflation. In other words, more openness leads to lower inflation.

Although our results are very general and cover general HSEMs, they are likely to be more useful in a setting with cross-sectional data because we are not allowing explicitly for some popular time series models for conditional heteroskedasticity. For example, we do not allow for GARCH type heteroskedasticity. Indeed, for some of the volatility models typically used in structural vector autoregressive analysis, no general tests for identification seem to be available. Developing such tests based on the ideas of the present paper may be worthwhile in future research.

A Appendix

A.1 Proof of Proposition 1

When any of the equations in (1) is multplied by -1, an observationally equivalent system results. Hence, any row in A_{01} can only be identified up to the scale ± 1. Accordingly, we define the true parameter point $\theta_0 = [a_1', \dots, a_r', \beta_{01}', \dots, \beta_{0r}']'$ only up to the scale ± 1 for $(a_1, \dots, a_r)$. In a compact neighborhood of θ_0, $\mathcal{N}_{\theta_0}$, under stated assumptions, we show that the uniform weak law of large numbers (WLLN) documented in Newey and McFadden (1994, Lemma 2.4) applies to $u_i u_i'$ and $\ln \sigma_{k,i}^2 + a_k' u_i u_i' a_k (\sigma_{k,i}^{-2} - 1)$ for $k = 1, \dots, r$. First, as each element in $u_i u_i'$ is bounded by $u_i' u_i$ and $\mathbb{E}(u_i' u_i) < \mathbb{E}[(u_i' u_i) \exp\{g(z_i)\}]$ is finite, the WLLN applies and $\hat{\Omega} \xrightarrow{p} B_0 B_0'$, where B_0 is $B = A^{-1}$ at the true parameter point.

Second,

$$\begin{aligned} |\ln \sigma_{k,i}^2 + a_k' u_i u_i' a_k(\sigma_{k,i}^{-2} - 1)| &\leq |F_k(z_i, \beta)| + |a_k' u_i|^2 (\exp\{g(z_i)\} + 1) \\ &\leq g(z_i) + m_a(u_i' u_i)(\exp\{g(z_i)\} + 1)|, \qquad k = 1, \ldots, r, \end{aligned}$$

where $m_a = \sup_{\mathcal{N}_{\theta_0}} \|a_k\|^2$ and the Chauchy-Schwarz inequality implies that $|a_k' u_i|^2 \leq \|a_k\|(u_i' u_i)$. Under A4, $\mathbb{E}[g(z_i) + m_a(u_i' u_i)(\exp\{g(z_i)\} + 1)|]$ is finite and independent of θ. It follows that the uniform WLLN holds, i.e., as functions of (a_k, β_k),

$$\ell_{k,n} \xrightarrow{p} \mathbb{E}(\ell_{k,n}) = -\mathbb{E}\big[\ln(\sigma_{k,i}^2) + a_k' B_0 H_{0i} B_0' a_k (\sigma_{k,i}^{-2} - 1)\big], \qquad k = 1, \ldots, r, \tag{14}$$

uniformly over $\mathcal{N}_{\theta_0}$ when $n \to \infty$. Here H_{0i} is H_i evaluated at θ_0.

We then consider the consistency of $(\hat{a}_1, \hat{\beta}_1)$. Given (14), we only need to show that $\mathbb{E}(\ell_{1,n})$ is uniquely maximized at θ_0. Clearly, for given $\sigma_{1,i}^2$ (or β_1), the quadratic form $a_1' B_0 H_{0i} B_0' a_1 (\sigma_{1,i}^{-2} - 1)$ in $(\ell_{1,i})$, subject to $a_1' B_0 B_0' a_1 = 1$, is minimized by the eigenvector a_1^* associated with the smallest eigenvalue μ_1^* in

$$B_0\big[\mathbb{E}\big(H_{0i}(\sigma_{1,i}^{-2} - 1)\big) - \mu_1 I_K\big] B_0' a_1 = 0. \tag{15}$$

Given that $H_{0i} = \text{diag}[\sigma_{01,i}^2, \ldots, \sigma_{0r,i}^2, 1, \ldots, 1]$, the eigenvalue is $\mu_1^* = \mathbb{E}(\sigma_{0k,i}^2(\sigma_{1,i}^{-2} - 1))$ for some $k \in \{1, \ldots, K\}$, where $\sigma_{0k,i}^2$ is $\sigma_{k,i}^2$ evaluated at θ_0 for $k = 1, \ldots, r$ and $\sigma_{0k,i}^2 = 1$ for $k = r+1, \ldots, K$. Then the concentrated objective function satisfies

$$\begin{aligned} \mathbb{E}(\ell_{1,n}) &= -\mathbb{E}\big[\ln(\sigma_{1,i}^2)\big] - \mu_1^* \\ &= -\mathbb{E}\big[\ln(\sigma_{1,i}^2) + \sigma_{0k,i}^2(\sigma_{1,i}^{-2} - 1)\big] \\ &\leq -\mathbb{E}\big[\ln(\sigma_{0k,i}^2) + \sigma_{0k,i}^2(\sigma_{0k,i}^{-2} - 1)\big] = -\mathbb{E}\ln(\sigma_{0k,i}^2), \end{aligned} \tag{16}$$

because the function $\ln(x) + x_0(x^{-1} - 1)$ is uniquely minimized at $x = x_0$. This result implies that the unique maximizer is $\sigma_{1,i}^2 = \sigma_{0k,i}^2$ for a $k \in \{1, \ldots, K\}$. Furthermore, as $\mathbb{E}\ln(\sigma_{0k,i}^2) < \ln(1)$ for $k \in \{1, \ldots, r\}$ by Jensen's inequality, the maximizer must be $\sigma_{1,i}^2 = \sigma_{0k_1,i}^2$ with

$$k_1 = \arg \min_{k \in \{1,\ldots,r\}} \mathbb{E}\ln(\sigma_{0k,i}^2). \tag{17}$$

Here, the maximizer $\sigma_{0k_1,i}^2$ is unique in the sense below. As $\mathbb{E}\ln(\sigma_{0k_1,i}^2) \leq \mathbb{E}\ln(\sigma_{0k,i}^2)$ for all $k \neq k_1$ and $\{\sigma_{01,i}^2, \ldots, \sigma_{0r,i}^2\}$ are linearly independent, Jensen's inequality leads to

$$\ln[\mathbb{E}(\sigma_{0k,i}^2/\sigma_{0k_1,i}^2)] > \mathbb{E}[\ln(\sigma_{0k,i}^2/\sigma_{0k_1,i}^2)] = \mathbb{E}\ln(\sigma_{0k,i}^2) - \mathbb{E}\ln(\sigma_{0k_1,i}^2) \geq 0,$$

i.e., $\mathbb{E}(\sigma_{0k,i}^2/\sigma_{0k_1,i}^2) > 1$ or $\mathbb{E}\big(\sigma_{0k,i}^2(\sigma_{0k_1,i}^{-2} - 1)\big) > 0$ for all $k \neq k_1$. Consequently, at the maximum, $\sigma_{1,i}^2 = \sigma_{0k_1,i}^2$, the only zero (smallest) element on the diagonal of $\mathbb{E}\big(H_{0i}(\sigma_{1,i}^{-2} - 1)\big)$

in (15) is at position k_1. It follows that $\mu_1^* = \mu_{0k_1} = 0$ and $B_0' a_1^* = \delta_1 e_K^{k_1}$, where $a_1^* = \delta_1 A_0' e_K^{k_1} = \delta_1 a_{0k_1}$ is the k_1^{th} column of A_0' up to the scale $\delta_1 = \pm 1$ and e_K^j is the j^{th} column of I_K. Given that $\mathbb{E}(\ell_{1,n})$ is continuous and uniquely maximized at $(\delta_1 a_{0k_1}, \beta_{0k_1})$, Theorem 2.1 of Newey and McFadden (1994) applies, i.e., $(\hat{a}_1, \hat{\beta}_1) \xrightarrow{p} (a_1^*, \beta_1^*) = (\delta_1 a_{0k_1}, \beta_{0k_1})$. The uniqueness of $(\delta_1 a_{0k_1}, \beta_{0k_1})$ implies identification.[3]

We use the same argument to show the consistency of $(\hat{a}_2, \hat{\beta}_2)$. To implement the restriction $a_2' B_0 B_0' a_1 = 0$, we let $a_2 = Q_{02}\rho_2$, where Q_{02} is a $K \times (K-1)$ matrix such that the augmented matrix $[a_1^*, Q_{02}]$ contains the full set of eigenvectors of (15). Under the uniform WLLN, we have

$$\ell_{2,n} \xrightarrow{p} \mathbb{E}(\ell_{2,n}) = -\mathbb{E}\big[\ln(\sigma_{2,i}^2) + \rho_2' Q_{02}' B_0 H_{0i} B_0' Q_{02}\rho_2(\sigma_{2,i}^{-2} - 1)\big],$$

uniformly in $\mathcal{N}_{\theta_0}$ when $n \to \infty$. The quadratic form $\rho_2' Q_{02}' B_0 H_{0i} B_0' Q_{02}\rho_2(\sigma_{2,i}^{-2} - 1)$ is minimized subject to $\rho_2'\rho_2 = 1$ by the vector ρ_2^* associated with the smallest eigenvalue μ_2^* in

$$Q_{02}' B_0 \big[\mathbb{E}\big(H_{0i}(\sigma_{2,i}^{-2} - 1)\big) - \mu_2 I_K\big] B_0' Q_{02}\rho_2 = 0.$$

Clearly, $\mu_2^* = \mathbb{E}\big(\sigma_{0k,i}^2(\sigma_{2,i}^{-2} - 1)\big)$ for a $k \in \{1, \ldots, K\}$ and $k \neq k_1$. Then

$$\begin{aligned} \mathbb{E}(\ell_{2,n}) &= -\mathbb{E}\big[\ln(\sigma_{1,i}^2) + \sigma_{0k,i}^2(\sigma_{2,i}^{-2} - 1)\big] \\ &\leq -\mathbb{E}\big[\ln(\sigma_{0k,i}^2) + \sigma_{0k,i}^2(\sigma_{0k,i}^{-2} - 1)\big] = -\mathbb{E}\ln(\sigma_{0k,i}^2) \end{aligned} \tag{18}$$

implies that the maximizer is $\sigma_{2,i}^2 = \sigma_{0k,i}^2$ and the maximum is $-\mathbb{E}\ln(\sigma_{0k,i}^2)$. As $\mathbb{E}\ln(\sigma_{0k,i}^2) < \ln(1)$ for $k \in \{1, \ldots, r\}$, the maximizing k must be $k_2 = \arg\min_{k\in\{1,\ldots,r\}, k\neq k_1} \mathbb{E}\ln(\sigma_{0k,i}^2)$. Correspondingly, $\mu_2^* = \mu_{0k_2} = 0$, $B_0' Q_{02}\rho_2^* = \delta_2 e_{k_2}$ and $a_2^* = Q_{02}\rho_2^* = \delta_2 A_0' e_{k_2} = \delta_2 a_{0k_2}$, where $\delta_2 = \pm 1$. Hence $\mathbb{E}(\ell_{2,n})$ is uniquely maximized by $(a_2^*, \beta_2^*) = (\delta_2 a_{0k_2}, \beta_{0k_2})$ and $(\hat{a}_2, \hat{\beta}_2) \xrightarrow{p} (\delta_2 a_{0k_2}, \beta_{0k_2})$. Similarly, it follows that $(\hat{a}_j, \hat{\beta}_j) \xrightarrow{p} (\delta_j a_{0k_j}, \beta_{0k_j})$ for $j = 3, \ldots, r$, where $\delta_j = \pm 1$.

We now assess the impact of using $\hat{D}$ and $\hat{u}_i = y_i - \hat{D}x_i$ instead of D and $u_i = y_i - Dx_i$, respectively. Under A5, the central limit theorem (CLT) of McLeish (1974) applies to $\text{vec}(u_i x_i')$ via the Cramér-Wold device. The OLS estimator satisfies $\hat{D} = D_0 + (\sum_{i=1}^n u_i x_i')(\sum_{i=1}^n x_i x_i')^{-1} = D_0 + O_p(n^{-1/2})$. As $\hat{u}_i = u_i + \dot{D}x_i$ with $\dot{D} = (D_0 - \hat{D}) = O_p(n^{-1/2})$, we find

$$\frac{1}{n}\sum_{i=1}^n \hat{u}_i\hat{u}_i' = \frac{1}{n}\sum_{i=1}^n \big(u_i u_i' + \dot{D}x_i u_i' + u_i x_i'\dot{D}' + \dot{D}x_i x_i'\dot{D}'\big) = \frac{1}{n}\sum_{i=1}^n u_i u_i' + O_p(n^{-1})$$

[3] It is easy to see that this result will break down if some of $\{\sigma_{01,i}^2, \ldots, \sigma_{0r,i}^2\}$ are proportional, in which case there will be two or more non-zero elements in $B_0' a_1^*$. For example, if $\sigma_{01,i}^2 = \sigma_{02,i}^2$ and $\mathbb{E}\ln(\sigma_{01,i}^2) = \min_{k\in\{1,\ldots,r\}} \mathbb{E}\ln(\sigma_{0k,i}^2)$, then $(\hat{a}_1, \hat{\beta}_1) \xrightarrow{p} (a_1^*, \beta_1^*) = (\delta_1 a_{01} + \delta_2 a_{02}, \beta_{01})$, where $\delta_1^2 + \delta_2^2 = 1$.

because $\sum_{i=1}^{n} u_i x_i' = O_p(n^{1/2})$. Under A4, the elements in $\mathbb{E}|x_i x_i' \sigma_{k,i}^{-2}| \leq \mathbb{E}|x_i x_i' e^{g(z_i)}|$ are finite, where $|\cdot|$ signifies element-wise absolute values. As the CLT also applies to $\text{vec}(u_i x_i')\sigma_{k,i}^{-2}$, it follows that

$$\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i \hat{u}_i' \sigma_{k,i}^{-2} = \frac{1}{n}\sum_{i=1}^{n} \big(u_i u_i' + \dot{D} x_i u_i' + u_i x_i' \dot{D}' + \dot{D} x_i x_i' \dot{D}'\big)\sigma_{k,i}^{-2} = \frac{1}{n}\sum_{i=1}^{n} u_i u_i' \sigma_{k,i}^{-2} + O_p(n^{-1})$$

for $k = 1, \ldots, r$. Thus the feasible objective function is related to the "ideal" one via

$$\hat{\ell}_{k,n} = -\frac{1}{n}\sum_{i=1}^{n}\Big[\ln(\sigma_{k,i}^2) + a_k' \hat{u}_i \hat{u}_i' a_k (\sigma_{k,i}^{-2} - 1)\Big] = \ell_{k,n} + O_p(n^{-1}), \qquad k = 1, \ldots, r,$$

which holds uniformly over a compact neighborhood of θ_0 and implies $\hat{\ell}_{k,n} \xrightarrow{p} \mathbb{E}(\ell_{k,n})$. Hence, our consistency argument also applies to the maximizers of $\hat{\ell}_{k,n}$. □

A.2 The Structure of $\mathcal{D}_{r\perp}$

We explicitly present the $r^2 \times \frac{1}{2}r(r+1)$ duplication matrix as

$$\mathcal{D}_r = \begin{bmatrix} I_r & 0 & \cdots & 0 & 0 \\ E_r^{12} & I_{r,-(1:1)} & \cdots & 0 & 0 \\ E_r^{13} & E_{r,-(1:1)}^{23} & \ddots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ E_r^{1(r-1)} & E_{r,-(1:1)}^{2(r-1)} & \cdots & I_{r,-(1:r-2)} & 0 \\ E_r^{1r} & E_{r,-(1:1)}^{2r} & \cdots & E_{r,-(1:r-2)}^{(r-1)r} & I_{r,-(1:r-1)} \end{bmatrix},$$

where $I_{r,-(1:l)}$ is the identity matrix I_r with its $(1,\ldots,l)^{\text{th}}$ columns being removed, E_r^{jk} is the $r \times r$ matrix with 1 in the $(j,k)^{\text{th}}$ position and 0 elsewhere, $E_{r,-(1:l)}^{jk}$ is E_r^{jk} with its $(1,\ldots,l)^{\text{th}}$ columns being removed. Let M be a $r \times r$ symmetric matrix with the lower triangular part of its k^{th} column denoted by m_k^h, i.e., $\text{vech}(M) = [m_1^{h\prime}, m_2^{h\prime}, \ldots, m_r^{h\prime}]'$. It can be verified that the k^{th} block of $\text{vec}(M) = \mathcal{D}_r \text{vech}(M)$, or the k^{th} column of M, is

$$m_k = \sum_{j=1}^{k-1} E_{r,-(1:j-1)}^{jk} m_j^h + I_{r,-(1:k-1)} m_k^h, \qquad k = 1, \ldots, r,$$

where $E_{r,-(1:0)}^{jk}$ is defined to be E_r^{jk}. The elements of $\mathcal{D}_r$ are either 0 or 1, where r columns, namely the first, the $r+1^{\text{th}}$, the $r+(r-1)+1^{\text{th}}$, …, the $\frac{1}{2}r(r+1)^{\text{th}}$, have only one 1, while the remaining $r(r-1)/2$ columns have two ones. The matrix $\mathcal{D}_{r\perp}$ can be constructed as the matrix consisting of the $r(r-1)/2$ columns of $\mathcal{D}_r$ that have two ones and, in each column, one of the ones (say the second one) is turned to -1. For example, when $r = 2$ and 3,

$$\mathcal{D}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathcal{D}_{2\perp} = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \mathcal{D}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \mathcal{D}_{3\perp} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}.$$

In general, $\mathcal{D}_{r\perp}$ can be expressed as

$$\mathcal{D}_{r\perp} = \begin{bmatrix} I_{r,-(1:1)} & 0 & \cdots & 0 \\ -E^{12}_{r,-(1:1)} & I_{r,-(1:2)} & \cdots & 0 \\ -E^{13}_{r,-(1:1)} & -E^{23}_{r,-(1:2)} & \ddots & 0 \\ \vdots & \vdots & & \vdots \\ -E^{1(r-1)}_{r,-(1:1)} & -E^{2(r-1)}_{r,-(1:2)} & \cdots & I_{r,-(1:r-1)} \\ -E^{1r}_{r,-(1:1)} & -E^{2r}_{r,-(1:2)} & \cdots & -E^{(r-1)r}_{r,-(1:r-1)} \end{bmatrix}.$$

As the non-zero elements of any column are in different positions in $\mathcal{D}_{r\perp}$, it follows that $\mathcal{D}'_{r\perp}\mathcal{D}_{r\perp} = 2I_{r(r-1)/2}$. □

A.3 Invertibility of $\Phi'_{0\perp} J_0 \Phi_{0\perp}$

Lemma 1. *$\Phi'_{0\perp} J_0 \Phi_{0\perp}$ is block diagonal and positive definite when $J_{0,22}$ is invertible.* □

Proof of Lemma 1 We write

$$\Phi'_{0\perp} J_0 \Phi_{0\perp} = \Gamma = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} & \Gamma_{13} \\ \Gamma'_{12} & \Gamma_{22} & \Gamma_{23} \\ \Gamma'_{13} & \Gamma'_{23} & \Gamma_{33} \end{bmatrix},$$

and find

$$
\begin{aligned}
\Gamma_{11} &= \mathcal{D}'_{r\perp}(I_r \otimes A_{01})J_{0,11}(I_r \otimes A'_{01})\mathcal{D}_{r\perp} = \mathbb{E}\left(\mathcal{D}'_{r\perp}(\Lambda_{0i}^{-1} - I_r) \otimes \Lambda_{0i}\mathcal{D}_{r\perp}\right),\\
\Gamma_{12} &= \mathcal{D}'_{r\perp}(I_r \otimes A_{01})J_{0,11}(I_r \otimes A'_{02}) = \mathbb{E}\left(D'_{r\perp}(\Lambda_{0i}^{-1} - I_r) \otimes A_{01}B_0H_{0i}B'_0A'_{02}\right) = 0,\\
\Gamma_{13} &= \mathcal{D}'_{r\perp}(I_r \otimes A_{01})J_{0,12} = -\mathbb{E}\left(D'_{r\perp}(\Lambda_{0i}^{-1} \otimes A_{01}B_0H_{0i}B'_0)\mathrm{diag}(a_{01}f'_{01,i}, \ldots, a_{0r}f'_{0r,i})\right) = 0,\\
\Gamma_{22} &= (I_r \otimes A_{02})J_{0,11}(I_r \otimes A'_{02}) = \mathbb{E}\left((\Lambda_{0i}^{-1} - I_r) \otimes I_{K-r}\right),\\
\Gamma_{23} &= (I_r \otimes A_{02})J_{0,12} = -\mathbb{E}\left((\Lambda_{0i}^{-1} \otimes A_{02}B_0H_{0i}B'_0)\mathrm{diag}(a_{01}f'_{01,i}, \ldots, a_{0r}f'_{0r,i})\right) = 0,\\
\Gamma_{33} &= J_{0,22}.
\end{aligned}
$$

The above results are easy to verify except the expressions for Γ_{11} and Γ_{13}. The case for $r = 1$ is covered by the lower 2×2 block sub-matrix of Γ. Hence, for $r > 1$, we only show $\Gamma_{13} = 0$ and that Γ_{11} is a positive definite diagonal matrix. For Γ_{13}, it is easily verified (using the definition $B_0 = A_0^{-1}$) that

$$
\Gamma_{13} = \mathbb{E}\left(\mathcal{D}'_{r\perp}\mathrm{diag}[e_r^1, \ldots, e_r^r]\mathrm{diag}[f'_{01,i}, \ldots, f'_{0r,i}]\right),
$$

where e_r^k is the k^{th} column of I_r. The fact that $\mathrm{diag}[e_r^1, \ldots, e_r^r]$ consists of r columns of $\mathcal{D}_r$ implies $\Gamma_{13} = 0$. For Γ_{11}, noting that $\Lambda_i E^{jk}_{r,-(1:l)} = \sigma^2_{j,i}E^{jk}_{r,-(1:l)}$, the k^{th} block row of $\mathcal{D}'_{r\perp}(\Lambda_{0i}^{-1} - I_r) \otimes \Lambda_{0i}$ is given by

$$
[0,\ \ldots,\ 0,\ I'_{r,-(1:k)}g_{k,i}\Lambda_{0i},\ -E^{k(k+1)\prime}_{r,-(1:k)}g_{k+1,i}\sigma^2_{0k,i}, \ldots, -E^{kr\prime}_{r,-(1:k)}g_{r,i}\sigma^2_{0k,i}], \quad k = 1, \ldots, r-1,
$$

where there are $k-1$ zero blocks and $g_{k,i} = (\sigma^{-2}_{0k,i} - 1)$. Furthermore, the $(k,k)^{\text{th}}$ and $(k,j)^{\text{th}}$ blocks of $\mathcal{D}'_{r\perp}(\Lambda_{0i}^{-1} - I_r) \otimes \Lambda_{0i}\mathcal{D}_{r\perp}$ are given, respectively, by

$$
\begin{aligned}
&I'_{r,-(1:k)}g_{k,i}\Lambda_{0i}I_{r,-(1:k)} + E^{k(k+1)\prime}_{r,-(1:k)}E^{k(k+1)}_{r,-(1:k)}g_{k+1,i}\sigma^2_{0k,i} + \cdots + E^{kr\prime}_{r,-(1:k)}E^{kr}_{r,-(1:k)}g_{r,i}\sigma^2_{0k,i}\\
&= g_{k,i}\Lambda_{0i,(k+1:r)} + E^{(k+1)(k+1)}_{r-k}g_{k+1,i}\sigma^2_{0k,i} + \cdots + E^{rr}_{r-k}g_{r,i}\sigma^2_{0k,i}\\
&= \mathrm{diag}(g_{k,i}\sigma^2_{0k+1,i} + g_{k+1,i}\sigma^2_{0k,i}, \ldots, g_{k,i}\sigma^2_{0r,i} + g_{r,i}\sigma^2_{0k,i})
\end{aligned}
$$

and

$$
\begin{aligned}
&I'_{r,-(1:k)}g_{k,i}\Lambda_{0i}E^{jk}_{r,-(1:j)} + E^{k(k+1)\prime}_{r,-(1:k)}E^{j(k+1)}_{r,-(1:j)}g_{k+1,i}\sigma^2_{0k,i} + \cdots + E^{kr\prime}_{r,-(1:k)}E^{jr}_{r,-(1:j)}g_{r,i}\sigma^2_{0k,i}\\
&= I'_{r,-(1:k)}g_{k,i}\sigma^2_{0j,i}E^{jk}_{r,-(1:j)} = 0 \quad \text{for} \quad j < k,
\end{aligned}
$$

where the fact that $E^{kl\prime}_{r,-(1:k)}E^{kl}_{r,-(1:k)} = E^{ll}_{r-k}$ and $E^{kl\prime}_{r,-(1:k)}E^{jl}_{r,-(1:j)} = 0$ for $l = k+1, \ldots, r$, is used and $\Lambda_{0i,(k+1:r)} \equiv \mathrm{diag}(\sigma^2_{0k+1,i}, \ldots, \sigma^2_{0r,i})$. It follows that $\Gamma_{11} = \mathbb{E}\,\mathrm{diag}(\gamma_1, \ldots, \gamma_{r-1})$ is diagonal, where

$$
\gamma_k = \mathrm{diag}(\underbrace{g_{k,i}\sigma^2_{0k+1,i} + g_{k+1,i}\sigma^2_{0k,i}, \ldots, g_{k,i}\sigma^2_{0r,i} + g_{r,i}\sigma^2_{0k,i}}_{r-k \text{ entries}})
$$

for $k = 1, \ldots, r-1$. A typical diagonal entry in Γ_{11} is

$$\mathbb{E}(g_{k,i}\sigma_{0j,i}^2 + g_{j,i}\sigma_{0k,i}^2) = \mathbb{E}\left(\frac{\sigma_{0j,i}^2}{\sigma_{0k,i}^2} + \frac{\sigma_{0k,i}^2}{\sigma_{0j,i}^2} - \sigma_{0j,i}^2 - \sigma_{0k,i}^2\right) = \mathbb{E}\left[\left(\frac{\sigma_{0j,i}}{\sigma_{0k,i}} - \frac{\sigma_{0k,i}}{\sigma_{0j,i}}\right)^2\right] > 0$$

because $\mathbb{E}(\sigma_{0j,i}^2) = 1$, and $\sigma_{0j,i}^2$ and $\sigma_{0k,i}^2$ are not proportional for $j \neq k$. This proves our claim. □

A.4 Proof of Proposition 2

We first show an intermediate result and then turn to the proof of Proposition 2.

Lemma 2. *Suppose that the assumptions of Proposition 1 hold. Assume further that A5 holds for $v_i = s_i(\theta_0)$. Then, $\sqrt{n}S_n(\theta_0) \xrightarrow{d} N(0, \Sigma_S)$, $\Sigma_S = \mathrm{var}\big(\sqrt{n}S_n(\theta_0)\big)$. This statement is also true when u_i in $s_i(\theta_0)$ is replaced by $\hat{u}_i$.*

Proof of Lemma 2 Denote $S_n = S_n(\theta_0)$ and $s_i = s_i(\theta_0)$. Under A5, the CLT of McLeish (1974) applies to $c's_i$. The Cramér-Wold device then implies that $\sqrt{n}S_n \xrightarrow{d} S \sim N(0, \Sigma_S)$. If A5 (iv) does not hold, i.e., there exists some constant vector p such that $n^{-1}\sum_{i=1}^n (p's_i)^2 \xrightarrow{p} 0$, let the space of all such p be spanned by a $K_\theta \times K_p$ matrix P, where $K_p < K_\theta$. Let $P_\perp$ be the orthogonal complement of P. As A5 (iv) holds for $P_\perp' s_i$, $\sqrt{n}P_\perp' S_n \xrightarrow{d} V \sim N(0, \Sigma_V)$. It follows that $\sqrt{n}S_n \xrightarrow{d} S \sim N(0, \Sigma_S)$ holds with $\Sigma_S = P_\perp(P_\perp' P_\perp)^{-1}\Sigma_V(P_\perp' P_\perp)^{-1}P_\perp'$. When u_i in s_i is replaced by $\hat{u}_i$, as shown at the end of the proof of Proposition 1, $S_n\big|_{\hat{u}_i} = S_n + O_p(n^{-1})$ because Λ_{0i} and $f_{0k,i}$ are bounded. The Lemma follows because $\sqrt{n}S_n\big|_{\hat{u}_i} = \sqrt{n}S_n + O_p(n^{-1/2})$. □

Proof of Proposition 2. The first-order conditions for maximizing (9) are

$$S_n(\hat{\theta}) - \Phi(\hat{\theta})\hat{\mu} = 0. \tag{19}$$

Taylor-expanding $S_n(\hat{\theta})$ and $\phi(\hat{\theta}) = 0$ at θ_0, we have

$$\begin{aligned} J_n(\bar{\theta})(\hat{\theta} - \theta_0) - \Phi(\hat{\theta})\hat{\mu} &= -S_n(\theta_0), \\ \Phi(\bar{\theta})'(\hat{\theta} - \theta_0) &= 0, \end{aligned} \tag{20}$$

where $\bar{\theta}$ is a point between θ_0 and $\hat{\theta}$. Denote $\Phi_0 = \Phi(\theta_0)$, $\bar{\Phi} = \Phi(\bar{\theta})$, $\hat{\Phi} = \Phi(\hat{\theta})$, $S_n = S_n(\theta_0)$, $\hat{S}_n = S_n(\hat{\theta})$, and $\bar{J}_n = J_n(\bar{\theta})$. As $\bar{\Phi}_\perp(\bar{\Phi}_\perp'\bar{\Phi}_\perp)^{-1}\bar{\Phi}_\perp' + \bar{\Phi}(\bar{\Phi}'\bar{\Phi})^{-1}\bar{\Phi}' = I_{K_\theta}$ and $(\hat{\theta} - \theta_0) = \bar{\Phi}_\perp(\bar{\Phi}_\perp'\bar{\Phi}_\perp)^{-1}\bar{\Phi}_\perp'(\hat{\theta} - \theta_0)$, the first equation in (20) can be written as

$$\bar{\Phi}_\perp'(\hat{\theta} - \theta_0) = (\bar{\Phi}_\perp'\bar{\Phi}_\perp)(\bar{\Phi}_\perp'\bar{J}_n\bar{\Phi}_\perp)^{-1}\bar{\Phi}_\perp'(\hat{\Phi}\hat{\mu} - S_n).$$

Solving the above equation together with the second equation in (20) gives

$$\begin{aligned}(\hat{\theta}-\theta_0) &= \bar{\Phi}_\perp(\bar{\Phi}_\perp'\bar{J}_n\bar{\Phi}_\perp)^{-1}\bar{\Phi}_\perp'(\hat{\Phi}\hat{\mu}-S_n)\\ &= \bar{\Phi}_\perp(\bar{\Phi}_\perp'\bar{J}_n\bar{\Phi}_\perp)^{-1}\bar{\Phi}_\perp'[\hat{\Phi}(\hat{\Phi}'\hat{\Phi})^{-1}\hat{\Phi}'\hat{S}_n-S_n]\end{aligned}$$

because, from (19), $\hat{\mu}=(\hat{\Phi}'\hat{\Phi})^{-1}\hat{\Phi}'\hat{S}_n$. Using $\hat{\Phi}(\hat{\Phi}'\hat{\Phi})^{-1}\hat{\Phi}'+\hat{\Phi}_\perp(\hat{\Phi}_\perp'\hat{\Phi}_\perp)^{-1}\hat{\Phi}_\perp'=I_{K_\theta}$ and $\hat{S}_n=S_n+\bar{J}_n(\hat{\theta}-\theta)$, we find

$$(\hat{\theta}-\theta_0)=-[I-\psi]^{-1}\bar{\Phi}_\perp(\bar{\Phi}_\perp'\bar{J}_n\bar{\Phi}_\perp)^{-1}\bar{\Phi}_\perp'[\hat{\Phi}_\perp(\hat{\Phi}_\perp'\hat{\Phi}_\perp)^{-1}\hat{\Phi}_\perp'S_n],$$

where $\psi=\bar{\Phi}_\perp(\bar{\Phi}_\perp'\bar{J}_n\bar{\Phi}_\perp)^{-1}\bar{\Phi}_\perp'[\hat{\Phi}(\hat{\Phi}'\hat{\Phi})^{-1}\hat{\Phi}'\bar{J}_n]$. As $n\to\infty$, $\bar{\Phi}\xrightarrow{p}\Phi_0$, $\hat{\Phi}\xrightarrow{p}\Phi_0$, $\bar{\Phi}_\perp\xrightarrow{p}\Phi_{0\perp}$, $\hat{\Phi}_\perp\xrightarrow{p}\Phi_{0\perp}$, $\psi\xrightarrow{p}0$, and $\bar{J}_n\xrightarrow{p}J_0$. Using Lemma 2 and the continuous mapping theorem, these results imply Proposition 2.

It remains to prove the alternative expression for Σ_θ below Proposition 2. If J_0 is invertible, the following result is implied by Lemma 3 below. Because Φ_0 is of full column rank, (20) can be solved for $(\hat{\theta}-\theta_0)$ and $\hat{\mu}$,

$$(\hat{\theta}-\theta_0)=-\big[\bar{J}_n^{-1}-\bar{J}_n^{-1}\hat{\Phi}(\bar{\Phi}'\bar{J}_n^{-1}\hat{\Phi})^{-1}\bar{\Phi}'\bar{J}_n^{-1}\big]S_n,\qquad \hat{\mu}=(\bar{\Phi}'\bar{J}_n^{-1}\hat{\Phi})^{-1}\bar{\Phi}'\bar{J}_n^{-1}S_n,$$

which lead to the alternative expression for Σ_θ below Proposition 2. □

Lemma 3. *Suppose that J is invertible. If U satisfy $\Phi_\perp'JU=\Phi_\perp'$ and $\Phi'U=0$, then $U=\Phi_\perp(\Phi_\perp'J\Phi_\perp)^{-1}\Phi_\perp'$.* □

Proof of Lemma 3 We only need to note that

$$\begin{bmatrix}\Phi_\perp'J\\ \Phi'\end{bmatrix}U=\begin{bmatrix}\Phi_\perp'\\ 0\end{bmatrix}\quad\text{and}\quad\begin{bmatrix}\Phi_\perp'J\\ \Phi'\end{bmatrix}^{-1}=[\Phi_\perp(\Phi_\perp'J\Phi_\perp)^{-1},J^{-1}\Phi(\Phi'J^{-1}\Phi)^{-1}].$$

□

A.5 Proof of Proposition 3

We again prove an intermediate result before turning to the proof of Proposition 3.

Lemma 4. *Suppose that the assumptions of Proposition 2 hold. Then the space spanned by the columns of $\hat{A}_2'$ is consitent for the space spanned by the columns of A_{02}' in the sense that*

$$A_{01}\Omega_0\hat{A}_2'\xrightarrow{p}0\qquad\textit{and}\qquad A_{02}\Omega_0\hat{A}_2'\xrightarrow{p}R,$$

where R is an orthogonal $(K-r)\times(K-r)$ matrix. This result also holds when Ω_0 is replaced by $\hat{\Omega}$. □

Proof of Lemma 4

As defined in Section 3, $\hat{A}_2' = Q_r[\hat{\rho}_2, \ldots, \hat{\rho}_{K-r+1}]$ consists of the eigenvectors associated with the $K - r$ largest eigenvalues in the system

$$(Q_r' \Psi_{r,n} Q_r - \mu I_{K-r+1})\rho = 0,$$

where $\Psi_{r,n} = n^{-1} \sum_{i-1}^{n} u_i u_i' (\hat{\sigma}_{r,i}^{-2} - 1)$ and $\hat{\sigma}_{r,i}^2$ is $\sigma_{r,i}^2$ evaluated at $\hat{\beta}_r$. These eigenvalues (being continuous functions of $\Psi_{r,n}$) converge in probability to the largest $K - r$ (positive) eigenvalues in

$$Q_{0r}' B_0 \Big[\mathbb{E}\big(H_{0i}(\sigma_{0r,i}^{-2} - 1)\big) - \mu I_K\Big] B_0' Q_{0r} \rho = 0.$$

The associated eigenvectors satisfy $B_0' Q_{0r}[\rho_2, \ldots, \rho_{K-r+1}] = [0, R]'$, for some orthogonal $(K - r) \times (K - r)$ matrix R. Hence the space spanned by the columns of $\hat{A}_2'$ converges in probability to the space spanned by $Q_{0r}[\rho_2, \ldots, \rho_{K-r+1}] = A_{02}' R$ in that $A_{01}\Omega_0 \hat{A}_2' \xrightarrow{p} 0$ and $A_{02}\Omega_0 \hat{A}_2' \xrightarrow{p} R$, although each column of $\hat{A}_2$ does not converge to a particular column of A_{02}'. The last statement of the lemma holds because $\hat{\Omega} \xrightarrow{p} \Omega_0$. □

Proof of Proposition 3

As $I_K = A_{02}' A_{02} \Omega_0 + A_{01}' A_{01} \Omega_0$, we may write

$$\hat{A}_2' = A_{02}' A_{02} \Omega_0 \hat{A}_2' + A_{01}' A_{01} \Omega_0 \hat{A}_2' \equiv A_{02}' \hat{d}_2 + A_{01}' \hat{d}_1.$$

Lemma 4 shows that $\hat{d}_2 \xrightarrow{p} R$ is invertible. This leads to $\hat{A}_2' \hat{d}_2^{-1} - A_{02}' = A_{01}' \hat{d}_1 \hat{d}_2^{-1}$. Since $0 = \hat{A}_1 \hat{\Omega} \hat{A}_2' = \hat{A}_1 \hat{\Omega} A_{02}' \hat{d}_2 + \hat{A}_1 \hat{\Omega} A_{01}' \hat{d}_1$, the Proposition 3 follows from

$$\hat{d}_1 \hat{d}_2^{-1} = -(\hat{A}_1 \hat{\Omega} A_{01}')^{-1} \hat{A}_1 \hat{\Omega} A_{02}' = -(\hat{A}_1 \hat{\Omega} A_{01}')^{-1} \Big[(\hat{A}_1 - A_{01})\hat{\Omega} + A_{01}(\hat{\Omega} - \Omega_0)\Big] A_{02}',$$

where $A_{01}\Omega_0 A_{02}' = 0$ is used, and the result that both $\sqrt{n}(\hat{A}_1 - A_{01})$ and $\sqrt{n}(\hat{\Omega} - \Omega_0)$ are asymptotically normal under the stated assumptions. This result also holds when u_i is replaced by $\hat{u}_i$ in computing $\hat{\Omega}$, as $\hat{\Omega}|_{\hat{u}_i} = \hat{\Omega} + O_p(n^{-1})$. □

A.6 Proof of Proposition 4

Proposition 4 holds because $\sqrt{n}\,\mathrm{vec}(\hat{C}_1' - C_{01}')$ can be expressed as

$$\sqrt{n}\,\mathrm{vec}(\hat{C}_1' - C_{01}') = \mathcal{T}\sqrt{n}[(\hat{\theta} - \theta_0)', \mathrm{vec}(\hat{D}' - D_0')']' + o_p(1).$$

Further, the asymptotic covariance matrix of $\sqrt{n}\,\mathrm{vec}(\hat{D}' - D_0')$ is clearly the probability limit of the covariance matrix of

$$\left[I_K \otimes \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\right] \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{vec}(x_i u_i'),$$

while the asymptotic covariance matrix of $\sqrt{n}(\hat{\theta} - \theta_0)$ is equivalent to the covariance matrix of

$$\Phi_{0\perp}(\Phi'_{0\perp} J_0 \Phi_{0\perp})^{-1}\Phi'_{0\perp}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i(\theta_0).$$

Hence, as $\text{vec}(x_i u_i') = u_i \otimes x_i$, the joint covariance matrix is found to be

$$\Sigma = \begin{bmatrix} \Sigma_\theta & \Sigma_{\theta D} \\ \Sigma'_{\theta D} & \Sigma_D \end{bmatrix},$$

where Σ_θ is defined in Proposition 2,

$$\Sigma_D = \Big[I_K \otimes \mathbb{E}(x_i x_i')^{-1}\Big]\mathbb{E}(u_i u_i' \otimes x_i x_i')\Big[I_K \otimes \mathbb{E}(x_i x_i')^{-1}\Big],$$

and

$$\Sigma_{\theta D} = \Phi_{0\perp}(\Phi'_{0\perp} J_0 \Phi_{0\perp})^{-1}\Phi'_{0\perp}\Big[\mathbb{E}\big(s_i(\hat{\theta})\text{vec}(x_i u_i')\big)\Big]\Big[I_K \otimes \mathbb{E}(x_i x_i')^{-1}\Big],$$

which lead to the estimator of Σ in (12). □

A.7 Proof of Proposition 5

Let $\hat{\beta}_k(a) = \arg\max_{\beta_k} \ell_{k,n}(\beta_k, a)$ and define

$$\mathcal{J}_n(\beta_k, a) = -\frac{\partial^2 \ell_{k,n}}{\partial\beta_k \partial\beta_k'} = \frac{1}{n}\sum_{i=1}^{n}\Big[\big(1 - \sigma_{k,i}^{-2} a' u_i u_i' a\big)\frac{\partial f_{k,i}}{\partial \beta_k'} + \sigma_{k,i}^{-2} a' u_i u_i' a f_{k,i} f_{k,i}'\Big],$$

where $a \in \Pi_k$. To use the distribution theory given by Hansen (1996) and Andrews and Ploberger (1995), we need to verify the following five results:

(i) $\hat{\beta}_k(a) \xrightarrow{p} \beta_{H_0}$ uniformly in Π_{0k};

(ii) $\mathcal{J}_n(\beta_k, a) \xrightarrow{p} \mathcal{J}_0(\beta_k, a)$ that is uniformly continuous in $\mathcal{N}_{\beta_{H_0}} \times \Pi_{0k}$;

(iii) $\mathcal{V}_n(\beta_k, a) \xrightarrow{p} \mathcal{V}_0(\beta_k, a)$ that is uniformly continuous in $\mathcal{N}_{\beta_{H_0}} \times \Pi_{0k}$;

(iv) $\mathcal{J}_0(\beta_{H_0}, a)$ and $\mathcal{V}_0(\beta_{H_0}, a)$ are uniformly positive definite in Π_{0k};

(v) $\sqrt{n}\mathcal{S}_n(\beta_{H_0}, a) \Rightarrow \mathcal{S}(a)$ on $a \in \Pi_{0k}$.

By B1, the uniform WLLN applies to $\ell_{k,n}$ under $\mathbb{H}_0$ (see Proof of Proposition 1), i.e.,

$$\begin{aligned} \ell_{k,n}(\beta_k, a) \xrightarrow{p} \mathbb{E}(\ell_{k,n}) &= -\mathbb{E}\big[\ln(\sigma_{k,i}^2) + a' B_0 H_{0i} B_0' a(\sigma_{k,i}^{-2} - 1)\big] \\ &= -\mathbb{E}\big[\ln(\sigma_{k,i}^2) + (\sigma_{k,i}^{-2} - 1)\big], \end{aligned}$$

uniformly in $\mathcal{N}_{\beta_{H_0}}$ for any $a \in \Pi_{0k}$. Clearly, $\mathbb{E}(\ell_{k,n})$ is uniquely maximized by $\sigma_{0k,i}^2 = 1$ or $\beta_k = \beta_{H_0}$. Hence $\hat{\beta}_k(a) \xrightarrow{p} \beta_{H_0}$ for any $a \in \Pi_{0k}$. Since Π_{0k} is compact, this convergence is uniform, i.e., $\sup_{a_k \in \Pi_k} \|\hat{\beta}_k(a) - \beta_{H_0}\| \xrightarrow{p} 0$, which verifies (i).

To verify (ii), (iii), and (iv), let $\mathcal{J}_{(i)}$ be the i^{th} summand of $\mathcal{J}_n(\beta_k, a)$, and similarly $\mathcal{V}_{(i)}$. It can be seen that both are bounded by quantities with finite means, because by B2,

$$\begin{aligned}
\|\mathcal{J}_{(i)}\| &= \left\| \left(1-\sigma_{k,i}^{-2} a' u_i u_i' a\right) \frac{\partial f_{k,i}}{\partial \beta_k'} + \sigma_{k,i}^{-2} a' u_i u_i' a f_{k,i} f_{k,i}' \right\|, \\
&\leq \left[1 + m_a u_i' u_i \exp(g(z_i))\right] g_2(z_i) + m_a u_i' u_i \exp(g(z_i)) g_1(z_i), \\
\|\mathcal{V}_{(i)}\| &= \left\| \left(1-\sigma_{k,i}^{-2} a' u_i u_i' a\right)^2 f_{k,i} f_{k,i}' \right\| \\
&\leq \left[1 + 2 m_a u_i' u_i \exp(g(z_i)) + m_a^2 (u_i' u_i)^2 \exp(2g(z_i))\right] g_1(z_i),
\end{aligned}$$

where $m_a = \max_{a \in \Pi_{0k}} \|a\|^2$. Hence, the uniform WLLN of Newey and McFadden (1994) applies:

$$\begin{aligned}
\mathcal{J}_n(\beta_k, a) \xrightarrow{p} \mathcal{J}_0(\beta_k, a) &= \mathbb{E}\Big[\left(1-\sigma_{k,i}^{-2} a' B_0 H_{0i} B_0' a\right) \frac{\partial f_{k,i}}{\partial \beta_k'} + \sigma_{k,i}^{-2} a' B_0 H_{0i} B_0' a f_{k,i} f_{k,i}'\Big] \\
&= \mathbb{E}\Big[\left(1-\sigma_{k,i}^{-2}\right) \frac{\partial f_{k,i}}{\partial \beta_k'} + \sigma_{k,i}^{-2} f_{k,i} f_{k,i}'\Big], \\
\mathcal{V}_n(\beta_k, a) \xrightarrow{p} \mathcal{V}_0(\beta_k, a) &= \mathbb{E}\Big[\left(1-\sigma_{k,i}^{-2} a' u_i u_i' a\right)^2 f_{k,i} f_{k,i}'\Big] \\
&= \mathbb{E}\Big[\Big(1 + \sigma_{k,i}^{-4} (\rho' \varepsilon_i^{(r_0+1:K)})^4 - 2\sigma_{k,i}^{-2} (\rho' \varepsilon_i^{(r_0+1:K)})^2\Big) f_{k,i} f_{k,i}'\Big],
\end{aligned}$$

uniformly in $\mathcal{N}_{\beta_{H_0}} \times \Pi_{0k}$, where $\rho'\rho = 1$. Clearly, $\mathcal{J}_0(\beta_k, a)$ and $\mathcal{V}_0(\beta_k, a)$ are uniformly continuous. $\mathcal{J}_0(\beta_{H_0}, a) = \mathbb{E}(f_{0k,i} f_{0k,i}')$ and $\mathcal{V}_0(\beta_{H_0}, a) = \mathbb{E}\big[\big((\rho' \varepsilon_i^{(r_0+1:K)})^4 - 1\big) f_{0k,i} f_{0k,i}'\big]$ are uniformly positive definite in Π_{0k} under B2.

To verify (v), we need to show that $\sqrt{n}\mathcal{S}_n(\beta_{H_0}, a)$ obeys the CLT for any $a \in \Pi_{0k}$ and that $\sqrt{n}\mathcal{S}_n(\beta_{H_0}, a)$ is stochastically equicontinuous in a. Under $\mathbb{H}_0$ ($\sigma_{0k}^2 = 1$) and at θ_0,

$$\begin{aligned}
\mathcal{S}_n(\beta_{H_0}, a) &= \frac{1}{n} \sum_{i=1}^n \left(1 - \rho' A_{02} u_i u_i' A_{02}' \rho\right) f_{0k,i} \\
&= \frac{1}{n} \sum_{i=1}^n \rho' \left(I_{K-r_0} - \varepsilon_i^{(r_0+1:K)} \varepsilon_i^{(r_0+1:K)\prime}\right) \rho f_{0k,i}
\end{aligned}$$

and $\mathbb{E}\,\mathcal{S}_n(\beta_{H_0}, a) = 0$ as $\mathbb{E}\big(\varepsilon_i^{(r_0+1:K)} \varepsilon_i^{(r_0+1:K)\prime} \big| W_i\big) = I_{K-r_0}$. The CLT of McLeish (1974) applies to $\sqrt{n}\mathcal{S}_n(\beta_{H_0}, a)$ for any $a \in \Pi_{0k}$ by B3. Let the matrix $\nu_i = [\nu_{jl,i}] = I_{K-r_0} - \varepsilon_i^{(r_0+1:K)} \varepsilon_i^{(r_0+1:K)\prime}$. For $a, b \in \Pi_{0k}$, we write $a = A_{02}'\rho$, $b = A_{02}'\varrho$ and

$$\mathcal{S}_n(\beta_{H_0}, b) - \mathcal{S}_n(\beta_{H_0}, a) = \sum_{j=1}^{K-r_0} \sum_{l=1}^{K-r_0} \frac{1}{n} \sum_{i=1}^n \nu_{jl,i} f_{0,i} (\varrho_j + \rho_j)(\varrho_l - \rho_l).$$

It follows that

$$\sup_{\|\varrho-\rho\|<\varphi} \left\| \sqrt{n}\big(\mathcal{S}_n(\beta_{H_0}, b) - \mathcal{S}_n(\beta_{H_0}, a)\big) \right\|$$

$$\leq \sup_{\|\varrho-\rho\|<\varphi} \sum_{j=1}^{K-r_0} \sum_{l=1}^{K-r_0} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nu_{jl,i} f_{0,i} \right\| |\varrho_j + \rho_j| \cdot |\varrho_l - \rho_l|$$

$$\leq \sum_{j=1}^{K-r_0} \sum_{l=1}^{K-r_0} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nu_{jl,i} f_{0,i} \right\| 2\varphi.$$

Then, for any $\tau > 0$ and $\zeta > 0$, there exists a $\varphi > 0$ such that

$$\limsup_{n\to\infty} P\left(\sup_{\|\varrho-\rho\|<\varphi} \left\| \sqrt{n}\big(\mathcal{S}_n(\beta_{H_0}, b) - \mathcal{S}_n(\beta_{H_0}, a)\big) \right\| > \zeta \right)$$

$$\leq \limsup_{n\to\infty} P\left(\sum_{j=1}^{K-r_0} \sum_{l=1}^{K-r_0} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nu_{jl,i} f_{0,i} \right\| > \frac{\zeta}{2\varphi} \right) < \tau,$$

as $n^{-1/2} \sum_{i=1}^{n} \nu_{jl,i} f_{0,i}$ converges in distribution and is hence uniformly tight. This verifies that $\sqrt{n}\mathcal{S}_n(\beta_{H_0}, a)$ is stochastically equicontinuous in Π_{0k} and, consequently, that $\sqrt{n}\mathcal{S}_n(\beta_{H_0}, a) \Rightarrow \mathcal{S}(a)$, a zero-mean Gaussian process on Π_{0k} (see Andrews (1994), p. 2251).

Furthermore, a Taylor expansion of $\mathcal{S}_n(\hat{\beta}_k(a), a)$ at β_{H_0} gives

$$\sqrt{n}(\hat{\beta}_k(a) - \beta_{H_0}) = -\mathcal{J}_n(\bar{\beta}_k(a), a)^{-1} \sqrt{n}\mathcal{S}_n(\beta_{H_0}, a) \Rightarrow -\mathcal{J}_0(\beta_{H_0}, a)^{-1}\mathcal{S}(a),$$

where $\bar{\beta}_k(a)$ is a point between $\hat{\beta}_k(a)$ and β_{H_0}. The Taylor expansion of $\ell_{k,n}\big(\hat{\beta}_k(a), a\big) - \ell_{k,n}(\beta_{H_0}, a)$ at $\hat{\beta}(a)$ leads to

$$n\ell_{k,n}(\hat{\beta}_k(a), a) = \frac{n}{2}\big(\hat{\beta}_k(a) - \beta_{H_0}\big)' \mathcal{J}_n\big(\bar{\beta}_k(a), a\big)\big(\hat{\beta}_k(a) - \beta_{H_0}\big) \Rightarrow \frac{1}{2}\mathcal{S}(a)'\mathcal{J}_0(\beta_{H_0}, a)^{-1}\mathcal{S}(a),$$

which delivers the results in (a)-(c) of Proposition 5. Moreover, Proposition 5 (d)-(e) follow from the continuous mapping theorem.

To show that the impact of using $\hat{u}_i$ instead of u_i is negligible, let $\hat{\mathcal{S}}_n$ be the version of $\mathcal{S}_n(\beta_k, a)$ using $\hat{u}_i$, similarly $\hat{\mathcal{J}}_n$ and $\hat{\mathcal{V}}_n$. B4 and the fact that $\hat{u}_i = u_i + (D_0 - \hat{D})x_i$ lead to $\|\sqrt{n}\big(\hat{\mathcal{S}}_n - \mathcal{S}_n(\beta_k, a)\big)\| = O_p(n^{-1/2})$, $\|\hat{\mathcal{J}}_n - \mathcal{J}_n(\beta_k, a)\| = O_p(n^{-1})$, and $\|\hat{\mathcal{V}}_n - \mathcal{V}_n(\beta_k, a)\| = O_p(n^{-1})$ uniformly over $\mathcal{N}_{\beta_{H_0}} \times \Pi_{0k}$, which concludes the proof of Proposition 5. □

A.8 Proof of Proposition 6

Under both $\mathbb{H}_0$ and $\mathbb{H}_1$, Proposition 3 implies

$$\hat{A}_2 u_i = (\hat{d}_2' A_{02} + \hat{d}_1' A_{01}) u_i = \hat{d}_2' \varepsilon_i^{(2)} + \hat{d}_1' \varepsilon_i^{(1)},$$

where $\hat{d}_2$ converges in probability to an orthogonal matrix R and $\hat{d}_1 = O_p(n^{-1/2})$. Then,

$$\hat{A}_2 u_i u_i' \hat{A}_2' = \hat{d}_2' \varepsilon_i^{(2)} \varepsilon_i^{(2)\prime} \hat{d}_2 + \hat{d}_2' \varepsilon_i^{(2)} \varepsilon_i^{(1)\prime} \hat{d}_1 + \hat{d}_1' \varepsilon_i^{(1)} \varepsilon_i^{(2)\prime} \hat{d}_2 + \hat{d}_1' \varepsilon_i^{(1)} \varepsilon_i^{(1)\prime} \hat{d}_1$$

holds, which leads to

$$\begin{aligned} \hat{\xi}_i &= \mathcal{D}_\tau^+ \mathrm{vec}(\hat{A}_2 u_i u_i' \hat{A}_2') \\ &= \mathcal{D}_\tau^+ (\hat{d}_2' \otimes \hat{d}_2') \mathcal{D}_\tau \xi_i \\ &\quad + \mathcal{D}_\tau^+ (\hat{d}_1' \otimes \hat{d}_2') \mathrm{vec}(\varepsilon_i^{(2)} \varepsilon_i^{(1)\prime}) + \mathcal{D}_\tau^+ (\hat{d}_2' \otimes \hat{d}_1') \mathrm{vec}(\varepsilon_i^{(1)} \varepsilon_i^{(2)\prime}) + \mathcal{D}_\tau^+ (\hat{d}_1' \otimes \hat{d}_1') \mathrm{vec}(\varepsilon_i^{(1)} \varepsilon_i^{(1)\prime}). \end{aligned}$$

It follows that, with $M_n = \mathcal{D}_\tau^+ (\hat{d}_2' \otimes \hat{d}_2') \mathcal{D}_\tau$ and $Z_i' = [1, w_i']$,

$$\sum_{i=1}^n \hat{\xi}_i Z_i' = M_n \sum_{i=1}^n \xi_i Z_i' + O_p(1)$$

since $\mathbb{E}(\varepsilon_i^{(2)} \varepsilon_i^{(1)\prime} \big| W_i) = 0$ and the CLT applies to $\mathrm{vec}\big(\mathrm{vec}(\varepsilon_i^{(1)} \varepsilon_i^{(2)\prime}) Z_i'\big)$. Under both $\mathbb{H}_0$ and $\mathbb{H}_1$, the feasible OLS estimator can then be expressed as

$$[\hat{\alpha}_0, \hat{\alpha}_1] = M_n \Big(\sum_{i=1}^n \xi_i Z_i' \Big) \Big(\sum_{i=1}^n Z_i Z_i' \Big)^{-1} + O_p(n^{-1}). \tag{21}$$

Under $\mathbb{H}_0$,

$$[\hat{\alpha}_0, \hat{\alpha}_1] - M_n[\alpha_0, 0] = M_n \Big(\sum_{i=1}^n \zeta_i Z_i' \Big) \Big(\sum_{i=1}^n Z_i Z_i' \Big)^{-1} + O_p(n^{-1}).$$

Because the CLT applies to $\zeta_i Z_i'$, the asymptotic distribution in Proposition 6 (a) is verified, i.e.,

$$\sqrt{n}\, \mathrm{vec}\Big([\hat{\alpha}_0, \hat{\alpha}_1] - M_n[\alpha_0, 0] \Big) \xrightarrow{d} N(0, V_\alpha),$$

where $V_\alpha = V_Z^{-1} \otimes M_0 V_\zeta M_0'$, $V_\zeta = \mathrm{var}(\zeta_i), V_Z = \mathbb{E}(Z_i Z_i')$, and M_0 is the probability limit of M_n. Since $\hat{\xi}_i - M_n \xi_i = O_p(n^{-1/2})$, $\hat{\zeta}_i = M_n \zeta_i + (\hat{\xi}_i - M_n \xi_i) + (M_n \alpha - \hat{\alpha}) Z_i$, and $\hat{\zeta}_i \hat{\zeta}_i' = M_n \zeta_i \zeta_i' M_n' + O_p(n^{-1/2})$, we find

$$\hat{V}_\alpha = \Big(\frac{1}{n} \sum_{i=1}^n Z_i Z_i' \Big)^{-1} \otimes \Big(\frac{1}{n} \sum_{i=1}^n \hat{\zeta}_i \hat{\zeta}_i' \Big) \xrightarrow{p} V_Z^{-1} \otimes M_0 V_\zeta M_0'$$

as claimed in Proposition 6 (b). The results in Proposition 6 (c) follow directly from (a) and (b). Under $\mathbb{H}_1$, applying the WLLN to (21) implies the result in Proposition 6 (d), where $\alpha = \mathbb{E}(\xi_i Z_i') \mathbb{E}(Z_i Z_i')^{-1}$. Under $\mathbb{H}_1$, the residual can be written as

$$\hat{\zeta}_i = M_n(\xi_i - \alpha Z_i) + (\hat{\xi}_i - M_n \xi_i) + (M_n \alpha - \hat{\alpha}) Z_i = M_n(\xi_i - \alpha Z_i) + o_p(1).$$

Using the WLLN and noting the values of α_0 and α_1, we verify Proposition 6 (e),

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\zeta}_i\hat{\zeta}_i' = \frac{1}{n}\sum_{i=1}^{n} M_n(\xi_i - \alpha Z_i)(\xi_i - \alpha Z_i)' M_n' + o_p(1) \xrightarrow{p} M_0(V_\xi - C_{\xi,w}V_w^{-1}C_{\xi,w}')M_0'.$$

The results in Proposition 6 (f) follow from (d) and that the matrix in Proposition 6 (e) is invertible.

To assess the effect of using $\hat{u}_i$, let $\bar{\xi}_i = \text{vech}(\hat{A}_2\hat{u}_i\hat{u}_i'\hat{A}_2')$. It holds that

$$\begin{aligned}\bar{\xi}_i &= \hat{\xi}_i + \mathcal{D}_\tau^+(\hat{A}_2\dot{D} \otimes \hat{A}_2)\text{vec}(u_i x_i') + \mathcal{D}_\tau^+(\hat{A}_2 \otimes \hat{A}_2\dot{D})\text{vec}(x_i u_i') \\ &\quad + \mathcal{D}_\tau^+(\hat{A}_2\dot{D} \otimes \hat{A}_2\dot{D})\text{vec}(x_i x_i'),\end{aligned}$$

where $\dot{D} = D_0 - \hat{D} = O_p(n^{-1/2})$. Because a CLT applies to $\text{vec}\big(\text{vec}(u_i x_i')Z_i'\big)$ and a WLLN applies to $\text{vec}(x_i x_i')Z_i'$, we have

$$\sum_{i=1}^{n}\bar{\xi}_i Z_i' = \sum_{i=1}^{n}\hat{\xi}_i Z_i' + O_p(1) = M_n\sum_{i=1}^{n}\xi_i Z_i' + O_p(1),$$

i.e., $\sum_{i=1}^{n}\bar{\xi}_i Z_i'$ is asymptotically equivalent to $\sum_{i=1}^{n}\hat{\xi}_i Z_i'$, which proves the last statement of Proposition 6. □

References

Andrews, D. W. K. (1994). Empirical process methods in econometrics, *in* R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, Elsevier, pp. 2247–2294.

Andrews, D. W. K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* **62**: 1383–1414.

Andrews, D. W. K. and Ploberger, W. (1995). Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative, *Annals of Statistics* **23**: 1609–1629.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2013). Identification-robust estimation and testing of the zero-beta CAPM, *Review of Economic Studies* **80**: 892–924.

Choi, I. and Phillips, P. C. (1992). Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations, *Journal of Econometrics* **51**(1): 113–150.

Cramér, H. and Wold, H. (1936). Some theorems on distribution functions, *Journal of the London Mathematical Society* **1**(4): 290–294.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* **64**(2): 247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* **74**(1): 33–43.

Doko Tchatoka, F. and Dufour, J. M. (2014). Identification-robust inference for endogeneity parameters in linear structural models, *Econometrics Journal* **17**: 165–187.

Dufour, J.-M. (2003). Identification, weak instruments and statistical inference in econometrics, *Canadian Journal of Economics* **36**: 767–808.

Farré, L., Klein, R. and Vella, F. (2013). A parametric control function approach to estimating the returns to schooling in the absence of exclusion restrictions: an application to the NLSY, *Empirical Economics* **44**: 111–133.

Hansen, B. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* **64**: 413–430.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. and Lee, T.-C. (1985). *The Theory and Practice of Econometrics*, 2nd edn, John Wiley and Sons, New York.

Klein, R. and Vella, F. (2010). Estimating a class of triangular simultaneous equation models without exclusion restrictions, *Journal of Econometrics* **154**: 154–164.

Lanne, M. and Lütkepohl, H. (2008). Identifying monetary policy shocks via changes in volatility, *Journal of Money, Credit and Banking* **40**: 1131–1149.

Lanne, M. and Saikkonen, P. (2007). A multivariate generalized orthogonal factor GARCH model, *Journal of Business and Economic Statistics* **25**: 61–75.

Lewbel, A. (2012). Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models, *Journal of Business and Economic Statistics* **30**: 67–80.

Lütkepohl, H. and Milunovich, G. (2016). Testing for identification in SVAR-GARCH models, *Journal of Economic Dynamics and Control* **73**: 241–258.

McLeish, D. (1974). Dependent central limit theorems and invariance principles, *Annals of Probability* **2**: 620–628.

Milunovich, G. and Yang, M. (2013). Simultaneous equation systems with heteroskedasticity: Identification, estimation, and stock price elasticities, *Research Paper 2013 ECON 01*, Australian School of Business, University of New South Wales.

Newey, W. K. and McFadden, D. L. (1994). Large sample estimation and hypothesis testing, *in* R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, Elsevier, pp. 2111–2245.

Phillips, P. C. B. (1989). Partially identified econometric models, *Econometric Theory* **5**: 181–240.

R-Team (2016). R: A language and environment for statistical computing.

Rigobon, R. (2003). Identification through heteroskedasticity, *Review of Economics and Statistics* **85**: 777–792.

Romer, D. (1993). Openness and inflation: Theory and evidence, *Quarterly Journal of Economics* **108**: 869–903.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments, *Econometrica* **65**: 557–586.

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* **48**: 817–838.

Wright, P. (1928). *The Tariff on Animal and Vegetable Oils*, Macmillan, New York.

Yang, M. (2014). Normality of posterior distribution under misspecification and nonsmoothness, and Bayes factor for Davies' problem, *Econometric Reviews* **33**: 305–336.