Simultaneous Equation Systems with Heteroskedasticity: 
Identification, Estimation, and Stock Price Elasticities

George Milunovich
Department of Economics, Macquarie University
North Ryde, NSW 2109, Australia

Minxian Yang
School of Economics, the University of New South Wales
Sydney, NSW 2052, Australia

Abstract
We give a set of identifying conditions for simultaneous equation systems (SES) with heteroskedasticity in the framework of Gaussian quasi maximum likelihood (QML). Our conditions rely on the presence of heteroskedasticity in the data rather than identifying restrictions traditionally employed in the literature. The QML estimators are shown to be consistent and asymptotically normal. Monte Carlo experiments indicate that the QML performs well in finite samples in comparison to GMM, even when conditional variance is mildly misspecified. We analyse the relationships between traded stock prices and volumes in the framework of SES. Based on a sample of the Russell 3000 stocks, our estimation results provide new evidence against homogeneous valuations.

Keywords: Endogeneity, Multivariate Structural Models, Quasi Maximum Likelihood, Asymptotics, Demand and Supply for Equities, Stock Prices and Volumes

JEL classification: C30, C13, C50, G12, G14

First version: January 2013
This version: April 2014
1 Introduction

Parameter identification in simultaneous equations systems (SES) traditionally relies on restrictions imposed on the parameter space (see Rothenberg (1971) and Richmond (1974)). However, difficulties in finding plausible restrictions have limited the pertinence of SES in many applications, see e.g. Rigobon (2003) and our application in Section 5. To mitigate such difficulties, recent literature has exploited heteroskedasticity for identification purposes. Although heteroskedasticity in regression models has a long history in econometrics\(^1\), it was typically regarded as a nuisance parameter that needs to be dealt with. The ability of “unequal variances” to assist identification in the context of SES was, until recently, largely neglected.

Intuitive arguments for using heteroskedasticity to identify SES are given as early as Wright (1928). More recently, in the context of time series data, Sentana and Fiorentini (2001) consider heteroskedasticity-based identification of factor models. Rigobon (2003), Bacchiocchi (2011), and Lanne, Lütkepohl and Maciejowska (2010) explore regime changes in the error variance for identifying parameters in structural vector autoregressions (SVAR). With GMM, Prono (2013) analyses the identification and estimation of a bivariate triangular system, in which the error terms follow a GARCH process. Milunovich and Yang (2013) provide sufficient conditions to identify the parameters in general SVAR models with GARCH-type error terms. In the context of cross-sectional data, Klein and Vella (2010) and Lewbel (2012) consider the identification and estimation of bivariate heteroskedastic SES, using semiparametric methods and generalised methods of moment (GMM) respectively. Lewbel (2012) also considers a nonlinear extension to the bivariate SES and discusses the possibilities of dealing with three or more endogenous variables within the GMM framework.

In this article, we consider the Gaussian quasi maximum likelihood (QML) estimation of SES in the presence of heteroskedasticity. We contribute to the literature by: a) providing a set of sufficient conditions for local identification of the structural parameters in the SES; b) deriving the asymptotic properties of the QML estimators; c) comparing the finite-sample performances of the QML estimators and the GMM estimators in Monte Carlo experiments; d) examining the relationships between traded prices and volumes of Russell 3000 stocks within the SES and documenting new empirical evidence on stock price elasticities. These results are new and not available elsewhere in the existing literature. As our setup and method

---

\(^1\) Early papers include Park (1966), Goldfeld and Quandt (1965), Glejser (1969), Harvey (1979), Breusch and Pagan (1979) and White (1980) among others. The goal of this literature is mainly about testing for the presence of heteroskedasticity or improving the estimation efficiency in a single equation context.
differ from the existing literature in several important aspects, we discuss our results in comparison with the existing literature next.

First, we study the identification and estimation of SES in the QML\(^2\) framework (also known as full information maximum likelihood in this context), whereas the existing literature mainly focuses on GMM (Lewbel (2012) and Prono (2013)) and control function (CF) based methods (Klein and Vella (2010)). Given that QML is an important inference tool in practice, the results of this article are valuable to practitioners. As identification conditions are related to the objective function used in estimation, the results from GMM and CF-based methods are not applicable to the QML framework. While Milunovich and Yang (2013) examine the QML for SVARs with GARCH type innovations, their results rely on specific GARCH functions and do not carry over to the general setup of this article, which applies to both time series and cross sectional data. The framework used by Bacchiocchi (2011), who considers QML for SES where heteroskedasticity is caused by structural changes in the variances of error terms (c.f. Rigobon (2003)), is also less general than ours. In fact, our results cover the model of Bacchiocchi (2011), see comments at the end of Section 2.1. Lanne et al (2010) discuss the identification of a SVAR (with QML) where the error term variances are Markov-switching but do not provide formal conditions for identification.

Second, we adopt the normalisation rule that the error terms in SES are uncorrelated, whereas Klein and Vela (2010) and Prono (2013) normalise their bivariate systems as triangular (i.e., the first endogenous variable does not appear in the second equation) and Lewbel (2012) imposes the normalisation that the covariances of exogenous variables and the product of the two error terms are zero. Complementary to these two types of normalisations, our normalisation rule is targeted toward applications where uncorrelated error terms are desirable. For instance, SVAR models typically require uncorrelated error terms. Cross-sectional demand-supply systems are another example in which uncorrelated error terms are preferred such that they can be interpreted as orthogonal shocks to demand and supply.

Third, our results provide new insights into the approach of heteroskedasticity based identification (HBI). It is known that a key condition for HBI, even under different normalisation rules, is that the conditional variances of different error terms are linearly independent (i.e., their non-zero linear combinations are non-zero), see Sentana and Fiorentini (2001) and Klein and Vela (2010) among others. We find that the identification

\(^2\)In the scenario of ML (i.e. the normality holds), the structural parameters in the SES cannot be identified with any other methods if they are not identifiable with the ML method (see Davidson (2002), p273).
with QML is achievable either when the conditional variances are linearly independent, or when the gradients (w.r.t. parameters) of log conditional variances are invertible, see Theorem 1 and Corollary 1 in Section 2.1 and their generalisation in Section 2.2. This result reveals that HBI is achievable even when the conditional variances of SES are linearly dependent, a finding that has not been recognised in the existing literature.

Fourth, our results are provided for general \( p \)-dimensional SES \((p \geq 2)\) systems, whereas the existing results from GMM and CF-based methods are primarily given for bivariate systems. Although Lewbel (2012) outlines a GMM procedure for systems with three or more endogenous variables, formal systematic generalisations are not provided in the existing literature for GMM and CF-based methods. Further, as the QML produces the estimates of all parameters in the SES (including parameters in the conditional variances), our results are particularly useful for applications where the heteroskedasticity itself is of interest. Although Klein and Vela (2010) use a semiparametric approach to estimate the parameters in conditional variances, Farre, Klein and Vela (2013) recommend that their semiparametric conditional variances be replaced by parametric exponential functions to ease implementation difficulties in practice. On the other hand, the GMM of Lewbel (2012) does not require the specifications of the conditional variances, and is suitable for applications that only focus on the mean-equation parameters.

We carry out Monte Carlo experiments to examine the performances of QML and GMM in finite samples, as well as the robustness of QML to misspecifications in the functional forms of heteroskedasticity. The data are generated with \( \chi^2 \)-distributed error terms such that the QML approach is truly “quasi”. We use two alternative specifications for the conditional variances to generate the data: a linear form and an exponential form, which may or may not match the conditional variances specified for the QML estimation. This design allows us to gauge the effect of misspecified conditional variances. The simulation results indicate that the QML approach estimates the structural parameters well and outperforms the GMM of Lewbel (2012) in terms of estimation biases and root mean squared errors (RMSE). The QML estimators also exhibit good performance in estimating the mean equation parameters in the cases where the conditional variances are mildly misspecified.

Finally, as an empirical application of our method, we estimate demand and supply elasticities for traded stocks. Standard finance theories predict that, in the absence of new information, investors may buy and sell any quantity of shares of a company without materially impacting its share price. In other words, as investors are assumed to have the
same (homogeneous) valuations, the supply and demand of stocks are predicted to be perfectly elastic. However, there is growing empirical evidence that suggests otherwise. Downward-sloping demand curves are found, amongst others, by Shleifer (1986), Kaul et al. (2000), Wurgler, and Zhuravskaya (2002), and Petajisto (2009, 2011), while evidence of upward-sloping supply schedules is found in Bradley et al. (1988), Bagwell (1992), Brown and Ryngaert (1992), and Hodrick (1999). These papers typically rely on event studies, which investigate corporate events that affect either demand or supply of an asset without altering its fundamental value. In contrast to the above literature, we model traded prices and volumes in a SES, in which the uncorrelatedness of error terms is required for interpretation, and use the QML to directly estimate the relationship between traded prices and volumes from a sample of companies included in the Russell 3000 index. After controlling for stock specific characteristics such as earnings, market beta, and book to market ratio, we find statistically significant demand and supply elasticities (−1.149 and 2.347 respectively), which provide new evidence against homogenous valuations. This application is the first study that directly estimates a demand-supply system for equities (to our knowledge).

The rest of the paper is organized as follows. In section 2, we present the model and identification conditions for bivariate and multivariate systems respectively. Section 3 contains the asymptotic properties of the QML estimators. Section 4 documents our simulation experiments. Section 5 is devoted to the empirical application. Major proofs and lemmas are collected in Section 6. Concluding remarks are in Section 7.

2 Model and Identification

2.1 Bi-variate Model

Consider the following bi-variate structural system that generates data

\[ y_{1i} = \beta_1^i x_i + \gamma_1 y_{2i} + \varepsilon_{1i}, \quad y_{2i} = \beta_2^i x_i + \gamma_2 y_{1i} + \varepsilon_{2i}, \quad i = 1, ..., n, \]

where \( i \) is the observation index (indicating either an individual or a time period), \( n \) is the sample size, \( x_i \) is a vector of exogenous or predetermined variables (including the constant one), while \( y_{1i} \) and \( y_{2i} \) are endogenous variables. Let \( W_i \) be the set of all exogenous or predetermined variables for index \( i \) that are observable. In the context of time series, \( W_i \) is the information set available at the end of period \( i - 1 \). Note that \( x_i \) can be either the totality or a
subset of $W_i$. To examine the identification issue in the presence of heteroskedasticity, but without imposing any restrictions on $(\beta_1', \gamma_1, \beta_2', \gamma_2)$, we make the following assumptions.

**ASSUMPTION A.** For all $i = 1, \ldots, n$, the following statements hold.

A1. $D = 1 - \gamma_1 \gamma_2 \neq 0$.

A2. The conditional variances of the error terms are given by

$$g_{1i} = \text{var}(\varepsilon_{1i}|W_i) = \exp\{F_1(z_i, \alpha_1)\},$$

and

$$g_{2i} = \text{var}(\varepsilon_{2i}|W_i) = \exp\{F_2(z_i, \alpha_2)\},$$

where $z_i$ is a subset of $W_i$, which may or may not be the same as $x_i$, and may include the constant one. The function $F_k(z_i, \alpha_k)$ is twice continuously differentiable with respect to $\alpha_k$ for $k = 1, 2$.

A3 The errors $[\varepsilon_{1i}, \varepsilon_{2i}]'$ have zero conditional means $E([\varepsilon_{1i}, \varepsilon_{2i}]'|W_i) = 0$ and zero conditional covariance $\text{cov}(\varepsilon_{1i}, \varepsilon_{2i}|W_i) = 0$.

A4 The standardised errors $[u_{1i}, u_{2i}]' \equiv [\varepsilon_{1i}/\sqrt{g_{1i}}, \varepsilon_{2i}/\sqrt{g_{2i}}]'$ are uncorrelated with $[u_{1j}, u_{2j}]'$ for any $j \neq i$.

Among these assumptions, A1 is required to express the system in the reduced-form model. A2 specifies the heteroskedasticity in the error terms $\varepsilon_{1i}$ and $\varepsilon_{2i}$. The function form $F_k(z_i, \alpha_k)$ needs to be specified in practice. The zero conditional mean in A3 clarifies the meaning of “exogenous variables” in this article. A3 is a normalising assumption that rules out the correlation between the structural errors for the same $i$. In fact, if $\text{cov}(\varepsilon_{1i}, \varepsilon_{2i}|W_i) \neq 0$, one can always construct an observationally equivalent system by multiplying a non-zero constant $c$ to the second equation, adding the result to the first equation, and re-arranging terms. Hence, A3 allows one to focus on a normalised version of the system that can be locally identified. The normalisation also facilitates the interpretation of the structural errors in some applications. For instance, in a demand-supply system, the error in the demand equation is interpreted as the aggregate force that shifts the demand curve but not the supply curve. A4 is commonly used in structural models for either cross-sectional or time series data.

For the system defined by (1) and Assumption A, the true parameter point $\theta_0$ is the vector of parameter values under which the data are generated. Because the positions of $y_{1i}$ and $y_{2i}$ are symmetric, (1) can be alternatively represented by

---

3 Similarly, Lewbel (2012, Assumption A2) normalises the system such that the two error terms satisfy $\text{cov}(z_i, \varepsilon_{1i}\varepsilon_{2i}) = 0$. On the other hand, Klein and Vella (2010) normalise the system into a triangular structure with $\gamma_2 = 0$ but allow the two error terms to be correlated.
as long as both \( \gamma_1 \) and \( \gamma_2 \) are non-zero. In other words, there are two observationally-equivalent representations (or parameterisations) for the same system. Hence, there are two true parameter points. For this reason, the interpretation of each equation and its regression coefficients must rely on the underlying economics (e.g., how a demand equation differs from a supply equation). In what follows, without requiring further knowledge of the underlying economics, we focus on the local identification of the parameters in (1), i.e., the true parameter point of (2) will be regarded as a remote point to that of (1).\(^4\)

If the distribution of \( (y_{1i}, y_{2i}|W_i) \) is known, the log likelihood has the form

\[
\sum_{i=1}^{n} \text{pdf}(y_{1i}, y_{2i}|W_i)
\]

and the maximum likelihood is the best inference method. As the distribution of \( (y_{1i}, y_{2i}|W_i) \) is generally unknown in practice, we use the Gaussian QML estimators that maximise the quasi log likelihood

\[
L_n(\theta) = -n \ln(2\pi) + n \ln(|D|) - \frac{1}{2} \sum_{i=1}^{n} \left[ \ln(g_{1i}^2g_{2i}^2) + \frac{\varepsilon_{1i}^2(\theta)}{g_{1i}^2} + \frac{\varepsilon_{2i}^2(\theta)}{g_{2i}^2} \right],
\]

where \( \varepsilon_{1i}(\theta) = y_{1i} - \beta_1'x_i - \gamma_1y_{2i} \) and \( \varepsilon_{2i}(\theta) = y_{2i} - \beta_2'x_i - \gamma_2y_{1i} \), \( D = 1 - \gamma_1\gamma_2 \) and \( \theta = [\gamma_1, \gamma_2, \beta_1', \beta_2', \alpha_1', \alpha_2']' \). At the true parameter point \( \theta_0 \), \( \varepsilon_{ki}(\theta_0) = \varepsilon_{ki} \) for \( k = 1, 2 \). In what follows, we denote \( \nabla_a = \partial L_n(\theta)/\partial a \), \( \nabla_{ab} = \partial^2 L_n(\theta)/\partial a \partial b \) and \( J_{ab} = -E(\nabla_{ab})/n \), where \( a \) and \( b \) can be \( \theta \) or a part of \( \theta \). Denote \( f_{ki} = \partial F_k(z_i, \alpha_k)/\partial \alpha_k \), \( h_k = E(\sum_{i=1}^{n} f_{ki}/n) \) and \( H_{kf} = E(\sum_{i=1}^{n} f_{ki}f_{ki}/n) \) and \( H_{kx} = E(\sum_{i=1}^{n} (x_i'x_i'/g_{ki})/n) \) for \( k = 1, 2 \).

The Taylor expansion of \( L_n(\theta) \) at \( \theta_0 \) suggests that \( \theta_0 \) is locally identifiable if and only if the quasi Fisher information matrix \( J_{\theta\theta} \) is positive definite in a neighbourhood of \( \theta_0 \) (see Gourieroux and Monfort (1989), section 3.4.4). Clearly, if the gradient \( \nabla_{\theta} \) of (4) is viewed as the moment vector for GMM, the QML estimator can also be interpreted as a GMM estimator. In GMM, the point \( \theta_0 \) is locally identifiable if and only if \( J_{\theta\theta} = -E(\partial \nabla_{\theta}/\partial \theta') \) is of full column rank in a neighbourhood of \( \theta_0 \) (see Newey and McFadden (1994), p2127). The main result is summarised in Theorem 1 below.

**THEOREM 1.** Let Assumption A hold. Assume further that \( H_{kx} \) and \( H_{kf} \) are of full rank for \( k = 1, 2 \) in a neighbourhood of \( \theta_0 \). Then, the true parameter point \( \theta_0 \) of (1) is locally identifiable if and only if

\(^4\)Our focus on the local identification is equivalent to making Assumption A4 of Lewbel (2011), which rules out one of (1) and (2) by assuming that if \( [\gamma_1, \gamma_2] \) is in the parameter space then \( [1/\gamma_2, 1/\gamma_1] \) is not.
where \( \lambda_{21} = \mathbb{E}(\sum_{i=1}^{n} (g_{2i}/g_{1i})/n) \) and \( \lambda_{12} = \mathbb{E}(\sum_{i=1}^{n} (g_{1i}/g_{2i})/n) \). Alternatively, (5) does not hold if and only if (i) \( \gamma_2^2 (1 - h'_1 H_{1f}^{-1} h_1) = \gamma_1^2 (1 - h'_2 H_{2f}^{-1} h_2) = 0 \) and (ii) \( g_{2i} = c g_{1i} \) for a non-zero constant \( c \) for all \( i \).

**Proof.** See the appendix. ■

By Lemma 1 in the Appendix, \( \lambda_{21} \lambda_{12} \geq 1 \) always and the equality holds if and only if \( g_{2i} = c g_{1i} \). Note also that \( 1 - h'_k H_{kf}^{-1} h_k \geq 0 \) always since \( \mathbb{E}(\sum_{i=1}^{n} (1 f_{ki}' [1, f_{ki}']) / n) \) is positive semi-definite for \( k = 1, 2 \). The equality holds if and only if \( \mathbb{E}(\sum_{i=1}^{n} (1 f_{ki}' [1, f_{ki}']) / n) \) is of reduced rank. For instance, \( 1 - h'_k H_{kf}^{-1} h_k = 0 \) when \( f_{ki} \) contains a constant element or when a non-zero linear combination of \( f_{ki} \) is a constant. According to Theorem 1, \( \theta_0 \) is locally identifiable when either \( g_{2i} \) is not a proportion of \( g_{1i} \) or at least one of \( \gamma_2^2 (1 - h'_1 H_{1f}^{-1} h_1) \) and \( \gamma_1^2 (1 - h'_2 H_{2f}^{-1} h_2) \) is positive. These conditions can be checked in practice. For example, it is easily deduced that \( \theta_0 \) is locally identifiable when only one of \( \epsilon_{1i} \) and \( \epsilon_{2i} \) is heteroskedastic (as \( g_{2i}/g_{1i} \) cannot be a constant).

That \( g_{2i} \) is not proportional to \( g_{1i} \) is a similar to the identifying condition of Klein and Vella (2010), which requires the ratio of the conditional standard deviations to depend on exogenous variables. For the GMM approach that does not require the specifications of the conditional variances, the identifying condition of Lewbel (2012) is that the covariances of the exogenous variable vector with the two squared error terms are linearly independent. It can be seen that \( g_{2i}/g_{1i} \) being a constant also implies the failure of Lewbel’s identifying condition. However, the reverse is not necessarily true (see Section 4 for an example).

According to Theorem 1, in the case that \( g_{2i} \) is proportional to \( g_{1i} \), the local identification may still be achievable when \( (1 - h'_k H_{kf}^{-1} h_k) > 0 \) for either \( k = 1 \) or \( 2 \). To our knowledge, this identification finding is new and has not been documented in the existing literature. We discuss this possibility further in the corollary below.

**COROLLARY 1.** Under the assumptions of Theorem 1, \( \theta_0 \) is locally identifiable when (i) \( \Omega_k = \mathbb{E}[\sum_{i=1}^{n} (f_{ki} - h_k)(f_{ki} - h_k)' / n] \) is invertible and (ii) \( \gamma_k = 0 \) for either \( k = 1 \) or \( 2 \), where \( l_k = 2 \) if \( k = 1 \) and \( l_k = 1 \) if \( k = 2 \).

**Proof.** Note that \( H_{kf} = \mathbb{E}(\sum_{i=1}^{n} f_{ki} f_{ki}' / n) = \mathbb{E}[\sum_{i=1}^{n} (f_{ki} - h_k)(f_{ki} - h_k)' / n] + h_k h'_k = \Omega_k + h_k h'_k \). Using the equality \( H_{kf}^{-1} = \Omega_k^{-1} - \Omega_k^{-1} h_k (1 + h'_k \Omega_k^{-1} h_k)^{-1} h'_k \Omega_k^{-1} \) when \( \Omega_k \) is...
invertible, we find that \( h_k^t H_k^{-1} h_k = h_k^t \Omega_k^{-1} h_k(1 + h_k^t \Omega_k^{-1} h_k)^{-1} < 1 \), which leads to \( \gamma_k^2 (1 - h_k^t H_k^{-1} h_k) > 0 \). The claim follows as \( \lambda_{21} \lambda_{21} \geq 1 \) according to Lemma 1 in the appendix.

In the following corollary, we further consider two leading specifications of \( g_{ki} \): an exponential form in which \( F_k(z_i, \alpha_k) = \alpha_k^t z_i \) or \( g_{ki} = \exp(\alpha_k^t z_i) \), and a linear form where \( F_k(z_i, \alpha_k) = \ln(\alpha_k^t z_i) \) or \( g_{ki} = \alpha_k^t z_i \), where the first element of \( z_i \) is the constant one. Notation-wise, let \( \alpha_k = [\alpha_{k0}, \alpha_{k1}]' \) have the same dimension as \( z_i \) for \( k = 1,2 \), where \( \alpha_{k0} \) corresponds to the constant one in \( z_i \). For the linear form, \( z_i \) is a vector of non-negative exogenous variables (e.g. element-by-element squared \( x_i \)) and \( \alpha_{k0} > 0 \) and \( \alpha_{k1} \geq 0 \) for \( k = 1,2 \).

**COROLLARY 2.** Under the assumptions of Theorem 1, \( \theta_0 \) is locally identifiable if and only if (i) \( \alpha_{11} \neq \alpha_{21} \) for the exponential form \( g_{ki} = \exp(\alpha_k^t z_i) \); or (ii) \( \alpha_1 \) and \( \alpha_2 \) are linearly independent for the linear form \( g_{ki} = \alpha_k^t z_i \); for \( k = 1,2 \).

**Proof.** For the exponential form, \( F_k(z_i, \alpha_k = \alpha_k^t z_i, f_{ki} = z_i, H_k = E(\sum_{i=1}^n z_i z_i' / n) \) and \( h_k = E(\sum_{i=1}^n z_i / n) \) for \( k = 1,2 \). For this case, the statement (i) in Theorem 1 is true as \( z_i \) contains the constant one and \( E(\sum_{i=1}^n [1 f_{ki}]' [1 f_{ki}] / n) \) is of reduced rank. Hence, given that \( g_{2i}/g_{1i} = \exp((\alpha_2 - \alpha_1)' z_i) \), the local identification is achieved if only if \( \alpha_{11} \neq \alpha_{21} \). For the linear form, \( F_k(z_i, \alpha_k = \ln(\alpha_k^t z_i), f_{ki} = z_i/\alpha_k^t z_i \). For this case, the statement (i) in Theorem 1 is also true because \( \alpha_k^t f_{ki} = 1 \) and \( E(\sum_{i=1}^n [1 f_{ki}]' [1 f_{ki}] / n) \) is of reduced rank. Hence, given that \( g_{2i}/g_{1i} = \alpha_2^t z_i/\alpha_1^t z_i \), the local identification is achieved if and only if \( \alpha_1 \) and \( \alpha_2 \) are linearly independent.

Our results here cover the case considered by Rigobon (2003, Proposition 1), where the observation index set \( \mathbb{I} = \{1, ..., n\} \) is partitioned into \( \mathbb{I}_0 \) and \( \mathbb{I}_1 \), \( z_i = [1, d_i]' \), \( d_i = 0 \) if \( i \in \mathbb{I}_0 \) and \( d_i = 1 \) if \( i \in \mathbb{I}_1 \). This is a special form of heteroskedasticity: the conditional variances of the structural errors of the observations in \( \mathbb{I}_1 \) differ from those in \( \mathbb{I}_0 \), which is also known as structural break in the time series literature.

### 2.2 Multivariate Model

We extend the results of Section 2.1 to the \( p \)-dimensional simultaneous equation system

\[
\frac{\Gamma^t y_i = B x_i + \epsilon_i}{\Gamma = \begin{bmatrix} 1 & \cdots & -\gamma_{1p} \\
\vdots & \ddots & \vdots \\
-\gamma_{p1} & \cdots & 1 \end{bmatrix}}
\]
where $B$ is a $p \times p_x$ dimension matrix, $p_x$ is the dimension of $x_i$, which is a subset of $W_i$ (the set of observable exogenous or predetermined variables). Traditionally, without exploitation of heteroskedasticity, the parameter identification in (6) can only be achieved by restrictions on $B$ and $\Gamma$, which are usually stated as the well-known order and rank conditions (e.g., see Judge et al (1985), p577). In our context, except that the diagonal entries of $\Gamma$ are normalised to ones, there are no other restrictions on $B$ and $\Gamma$. We consider the identification of the parameters in (6) by exploiting the heteroskedasticity in the error terms. Assumption A is re-expressed in matrix form as follows.

**ASSUMPTION B.** For all $i = 1, \ldots, n$, the following statements hold.

B1. $|\Gamma| \neq 0$.

B2. The conditional variances of the error terms are given by

$$g_{ki} = \text{var}(\varepsilon_{ki}|W_i) = \exp\{F_k(z_i, \alpha_k)\} \quad \text{for} \quad k = 1, \ldots, p,$$

where $z_i$ is a subset of $W_i$, which may or may not be the same as $x_i$, and may include the constant one. The function $F_k(z_i, \alpha_k)$ is twice continuously differentiable with respect to $\alpha_k$.

B3. The error vector $\varepsilon_i = [\varepsilon_{1i}, \ldots, \varepsilon_{pi}]'$ satisfies $E(\varepsilon_i | W_i) = 0$ and $\text{var}(u_i | W_i) = G_i$ with $G_i = \text{diag}[g_{1i}, \ldots, g_{pi}]$.

B4. The standardised error vector $u_i \equiv G_i^{-1/2}\varepsilon_i$ is uncorrelated with $u_j$ for any $j \neq i$.

The remarks under Assumption A also apply here. In particular, the covariances of the elements in $\varepsilon_i$ are normalised to zero for identification purposes. Similar to (2) for the bi-variate system, there are alternative representations of (6). Focusing on the true parameter point $\theta_0$ of (6) for one particular representation and treating other representations as remote points, we consider the local identification of $\theta_0$. Hence the interpretation of each equation in (6) must rely on relevant economic theory. Again, the QML estimators maximise the log quasi likelihood

$$L_n(\theta) = \frac{n}{2} \ln(2\pi) + n \ln(|\Gamma|) - \frac{1}{2} \sum_{i=1}^n \ln(|G_i|) + \text{tr}[G_i^{-1}\varepsilon_i(\theta)\varepsilon_i(\theta)'],$$

where $\theta$ stands for the vector of free parameters in (6) and Assumption B, and $\varepsilon_i(\theta) = \Gamma y_i - Bx_i$. Using the same definitions for $h_k, H_{kf}$ and $H_{kx}$ as in Section 2.1 for $k = 1, \ldots, p$, we state the main result in the following theorem.
THEOREM 2. Let Assumption B hold. Assume further that $H_{kx}$ and $H_{kf}$ are of full rank for $k = 1, ..., p$ in a neighbourhood of the true parameter point $\theta_0$. Then, $\theta_0$ is locally identifiable if and only if

\begin{equation}
J_{\Gamma, B \alpha} = R_p \left( I_p \otimes \Gamma^{-1} \right) [R_p V^{-1} \psi V^{-1} \psi' + S_p \mu S_p'] (I_p \otimes \Gamma^{-1'}) R_p
\end{equation}

is positive definite, where $\mu = \text{diag}[\mu_1, ..., \mu_p]$, $\mu_k = 2(1 - h'_k H_{kj} h_k)$, $\psi = \text{diag}[\psi_1, ..., \psi_p]$, $\psi_k = \text{diag}[\lambda_{1k}, ..., \lambda_{k-1,k}, (\lambda_{k+1,k} - \lambda_{k,k+1}^{-1}), ..., (\lambda_{p,k} - \lambda_{k,p}^{-1})]$, $\lambda_{lk} = \frac{1}{n} \sum_{t=1}^{n} g_{lt} / g_{ki}$ for all $l, k = 1, ..., p$ and $k \neq l$, $R_p = \text{diag}[I_{p-1}, ..., I_{p-p}], I_{p-k}$ is the $p$-dimensional identity matrix $I_p$ with the $k$th column being deleted, $S_p = \text{diag}[e_1, ..., e_p]$, $e_k$ is the $k$th column of $I_p$,

\[ V = \begin{bmatrix}
I_{p-1} & v_{12} & \cdots & v_{1p} \\
0 & I_{p-1} & \cdots & v_{2p} \\
& \vdots & \ddots & \vdots \\
0 & 0 & \cdots & I_{p-1}
\end{bmatrix}, \quad v_{kl} = -e_{l-k} e'_k \lambda_{kl}^{-1}, \quad k, l = 1, ..., p, \quad k < l,
\]

and $e_{k-l}$ is $e_k$ with the $l$th entry being removed.

Proof. See the appendix. ■

Here the presence of the matrices $R_p$ and $S_p$ is caused by the normalisation that the diagonal entries in $\Gamma$ are ones (not free parameters). Clearly, $J_{\Gamma, B \alpha}$ in (8) is the sum of two positive semi-definite matrices because $\lambda_{jk} - \lambda_{kj}^{-1} \geq 0$ (by Lemma 1 in the appendix) and $\mu_k \geq 0$ always (see remarks under Theorem 1). A sufficient condition for the local identification is stated in the corollary below.

COROLLARY 3. Under the conditions of Theorem 2, the true parameter point $\theta_0$ is locally identifiable if the error conditional variance of each equation is not proportional to that of another equation, i.e., the ratio $g_{ji} / g_{ki}$ is not a constant for all $j, k = 1, ..., p$ and $k \neq j$.

Proof. When $\frac{g_{ji}}{g_{ki}}$ is not a constant, by Lemma 1 in the appendix, $\lambda_{jk} - \lambda_{kj}^{-1} > 0$. Hence, (8) is positive definite because the second term in (8) is positive semi-definite and the matrix $R_p' \left( I_p \otimes \Gamma^{-1} \right) R_p V^{-1}$ is of full rank. ■

In particular, Corollary 3 can hold for the case where the error of one of the equations in (6) is homoscedastic. Further, for the exponential form and the linear form of conditional variances, as indicated in Corollary 2, the condition “$g_{ji} / g_{ki}$ is not a constant” is sufficient as well as necessary for the local identification because $\mu_k = 0$ for all $k$ in these cases.
Similar to the 2-dimensional system in Section 2.1, it is possible that, even when \( g_{ji}/g_{kl} \) is constant for some \( k \neq j \), \( J_{\Gamma, Ba} \) in (8) can still be positive definite if \( \mu_l > 0 \) for certain suitable \( l \). To consider this possibility, we denote \( \mathbb{N}(M) \) as the null-space of a matrix \( M \), which is the vector space spanned by \( \{ \eta : M\eta = 0 \} \).

**COROLLARY 4.** Under the conditions of Theorem 2, the true parameter point \( \theta_0 \) is locally identifiable if and only if the set

\[
\mathbb{N}(\mu) \cap T[\mathbb{N}(\psi) - \{0\}]
\]

is empty, i.e., \( T\eta \notin \mathbb{N}(\mu) \) for non-zero \( \eta \in \mathbb{N}(\psi) \), where the transformation \( T \), from \( \mathbb{N}(\psi) \subset \mathbb{R}^{p(p-1)} \) to \( \mathbb{R}^p \), is defined by the matrix \( T = S'_p(I_p \otimes \Gamma^{-1})R_p[R'_p(I_p \otimes \Gamma^{-1})R_p]^{-1}V' \).

**Proof.** That (8) is not positive definite is equivalent to that there exists a non-zero vector \( a \) such that \( \psi V^{-1}R'_p(I_p \otimes \Gamma^{-1})R_pa = 0 \) and \( \mu S'_p(I_p \otimes \Gamma^{-1})R_pa = 0 \), as both \( \psi \) and \( \mu \) are diagonal and non-negative. The first term is zero if and only if \( \eta = V^{-1}R'_p(I_p \otimes \Gamma^{-1})R_pa \) is a non-zero element of \( \mathbb{N}(\psi) \). The second term is zero if and only if \( S'_p(I_p \otimes \Gamma^{-1})R_pa = T\eta \) is an element of \( \mathbb{N}(\mu) \). Hence, that (8) is not positive definite is equivalent to that there exist a non-zero \( \eta \in \mathbb{N}(\psi) \) such that \( T\eta \in \mathbb{N}(\mu) \), which in turn is equivalent to that (9) is non-empty.

This condition is checkable given that both \( \psi \) and \( \mu \) are diagonal. In fact, the dimension of \( \mathbb{N}(\psi) \) equals to the number of zeros in the diagonal entries of \( \psi \). The basis of \( \mathbb{N}(\psi) \) may be expressed as \( p(p-1) \)-dimensional unit vectors, each of which contains all zeros but one one at the position corresponding to a zero in the diagonal entries of \( \psi \). For each \( \eta \) in the basis of \( \mathbb{N}(\psi) \), it is easy (but can be tedious) to check whether or not \( T\eta \) is zero or its non-zero entries correspond to the zeros in the diagonal entries of \( \mu \). The mapping defined by the matrix \( T \) depends on the free parameters in \( \Gamma \) and \( \lambda_{kl} \). Note that a special case is when \( \mathbb{N}(\psi) - \{0\} \) itself is empty (or when \( \mathbb{N}(\psi) \) only contains the vector 0), which is covered by Corollary 3. The two examples below serve the purpose of demonstration.

**Example 1.** For the bivariate case \( (p = 2) \), we have

\[
\mu = \begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{bmatrix}, \quad \psi = \begin{bmatrix} \lambda_{21} - \lambda_{12}^{-1} & 0 \\ 0 & \lambda_{12} \end{bmatrix}, \quad V = \begin{bmatrix} 1 & -\lambda_{12}^{-1} \\ 0 & 1 \end{bmatrix}, \quad T = \begin{bmatrix} \gamma_2 & 0 \\ 0 & \gamma_1 \end{bmatrix}V'.
\]

When \( \lambda_{21} - \lambda_{12}^{-1} = 0 \), the basis of \( \mathbb{N}(\psi) \) is \( \eta_1 = [1,0]' \). In this case,

\[
T\eta_1 = \begin{bmatrix} \gamma_2 \\ -\gamma_1 \lambda_{12}^{-1} \end{bmatrix} \quad \text{and} \quad \eta_1'T\mu T\eta_1 = \gamma_2^2 \mu_1 + \lambda_{12}^{-2} \gamma_1^2 \mu_2.
\]
Hence $\theta_0$ is not identifiable if and only if both $\gamma_2^2 \mu_1 = 0$ and $\gamma_1^2 \mu_2 = 0$. Additionally,

$$J_{\Gamma, Ba} = \frac{1}{d} \begin{bmatrix} \lambda_{21} + \gamma_2^2 \mu_1 & 1 \\ \lambda_{12} + \gamma_1^2 \mu_2 \end{bmatrix},$$
on which the claim of Theorem 1 is based.

**Example 2.** For the trivariate case ($p = 3$), we have

$$\mu = \text{diag}[\mu_1, \mu_2, \mu_3], \quad \psi = \text{diag}[\psi_1, \psi_2, \psi_3],$$

$$\psi_1 = \begin{bmatrix} \lambda_{21} - \lambda_{12}^{-1} \\ 0 \\ \lambda_{31} - \lambda_{13}^{-1} \end{bmatrix}, \quad \psi_2 = \begin{bmatrix} \lambda_{12} \\ 0 \\ \lambda_{32} - \lambda_{23}^{-1} \end{bmatrix}, \quad \psi_3 = \begin{bmatrix} \lambda_{13} \\ 0 \\ \lambda_{23} \end{bmatrix},$$

$$V = \begin{bmatrix} l_2 & v_{12} & v_{13} \\ 0 & l_2 & v_{23} \end{bmatrix}, \quad v_{12} = \begin{bmatrix} \lambda_{12}^{-1} \\ 0 \\ 0 \end{bmatrix}, \quad v_{13} = \begin{bmatrix} 0 \\ -\lambda_{13}^{-1} \\ 0 \end{bmatrix}, \quad v_{23} = \begin{bmatrix} 0 \\ 0 \\ -\lambda_{23}^{-1} \end{bmatrix},$$

$$T = \begin{bmatrix} \bar{m}_{1,-1}M_1^{-1}' & 0 & 0 \\ 0 & \bar{m}_{2,-2}M_2^{-1}' & 0 \\ 0 & 0 & \bar{m}_{3,-3}M_3^{-1}' \end{bmatrix},$$

where $M_1 = [\bar{m}_{2,-1}, \bar{m}_{3,-1}]$, $M_2 = [\bar{m}_{1,-2}, \bar{m}_{3,-2}]$, $M_3 = [\bar{m}_{1,-3}, \bar{m}_{2,-3}]$, $\bar{m}_k$ is the $k$th column of $\Gamma^{-1}$ and $\bar{m}_{k,l}$ is $\bar{m}_k$ with the $l$th entry being removed. Alternatively, $M_k$ is simply $\Gamma^{-1}$ with the $k$th column and row being deleted. When $\lambda_{21} - \lambda_{12}^{-1} = 0$, the basis of $\mathbb{H}(\psi)$ is $\eta_1 = [1,0,0,0,0]'$. Then,

$$T\eta_1 = \left[ ([1,0]M_1^{-1}\bar{m}_{1,-1}, [-\lambda_{12}^{-1}, 0]M_2^{-1}\bar{m}_{2,-2}, 0)'ight]$$

and

$$\eta_1'T'\mu T\eta_1 = \left( [1,0]M_1^{-1}\bar{m}_{1,-1} \right)^2 \mu_1 + \left( [-\lambda_{12}^{-1}, 0]M_2^{-1}\bar{m}_{2,-2} \right)^2 \mu_2 .$$

In this case, $\theta_0$ is not identifiable if and only if both terms in the above equation are zero.

**3 Asymptotic Properties of QML Estimators**

Let the true parameter vector $\theta_0$ contain the free parameters in $(\Gamma_0, B_0, \alpha_0)$. Define the neighbourhood $\Theta_\delta = \{ \theta: \| \theta - \theta_0 \| \leq \delta \}$ of $\theta_0$ for a fixed $\delta > 0$, where $\| \cdot \|$ is the Euclidean norm. We consider the asymptotic properties of the QML estimator of $\theta_0$ under the following assumptions.

**ASSUMPTION C.**

**C1.** The size of $\Theta_\delta$, $\delta$, is chosen such that $\Gamma$ is invertible and $-E(\partial^2 \frac{1}{n} L_n(\theta)/\partial \theta \partial \theta')$ is positive definite in $\Theta_\delta$ under the conditions of Theorem 2.
The parameterisation of the model is such that \( \theta_0 \) is an interior point of \( \Theta_\delta \).

The data are either iid or ergodic stationary such that

(i) \( \text{E}[\ln(|G_i|)], \text{E}(G_i^{-1} \otimes x_i x_i'), \text{E}(G_i^{-1} \otimes G_i). \text{E}(\phi_i) \) and \( \text{E}(\phi_i \phi_i') \) are finite for every \( \theta \in \Theta_\delta \), where \( \phi_i' = \text{diag}[f_{1i}', ..., f_p'] \);

(ii) the weak law of large numbers (WLLN) holds for \( \ln(|G_i|), x_i \varepsilon_i' G_i^{-1}, x_i x_i' G_i^{-1} \) and \( \varepsilon_i \varepsilon_i' G_i^{-1} \) for every \( \theta \in \Theta_\delta \).

The central limit theorem (CLT) holds for

\[
\begin{bmatrix}
\text{vec}(x_i \varepsilon_i' G_{i0}^{-1}) \\
\text{vec}(\varepsilon_i \varepsilon_i' G_{i0}^{-1} - I_p) \\
[f_{1i0} (\varepsilon_{11i0}^2 - 1), ..., f'_{p0i} (\varepsilon_{p1i0}^2 - 1)]'
\end{bmatrix}
\]

where \( G_{i0}, g_{ki0} \) and \( f_{ki0} \) are \( G_i, g_{ki} \) and \( f_{ki} \) evaluated at \( \theta_0 \).

Given that \( J_{\theta \theta} = -\text{E}(\partial^2 \ln L_n(\theta_0)/\partial \theta \partial \theta') \) is positive definite under the assumptions of Theorem 2, the \( \delta \) required in C1 exists. Since the system in (6) has multiple different representations and corresponding true parameter points, we use C1 to focus on the neighbourhood of one particular true parameter point. For C2, we note that the parameters in the linear-form conditional variance could be on the boundary defined by positivity restrictions. Nonetheless, C2 usually holds as long as the conditional variance \( G_i \) is suitably parameterised (e.g., if \( \alpha_{ij} \) is required to be nonnegative, it is re-parameterised as \( \alpha_{ij} = \alpha_{ij}^2 \)).

Since the WLLN and CLT generally hold for independent or stationary data (e.g., see Davidson (1994) among others for primitive conditions on WLLN and CLT), C3 and C4 are mild requirements on the data and the conditional variance. Indeed, C3 and C4 are similar to the traditional requirements on the generalised or weighted least squares for the reduced-form system of (6). The asymptotic properties are summarised in the theorem below, which is proved in the appendix.

**THEOREM 3.** Let the assumptions in Theorem 2 and Assumption C hold. Then, the QML estimator \( \hat{\theta}_n = \arg \max_{\theta \in \Theta_\delta} L_n(\theta) \) is a consistent estimator of \( \theta_0 \) and is asymptotically normally distributed with \( \sqrt{n}(\hat{\theta}_n - \theta_0) \overset{d}{\rightarrow} N(0, J_{\theta \theta}^{-1} \Omega J_{\theta \theta}^{-1}) \), where \( \Omega = \text{var}(\partial \ell_i(\theta_0)/\partial \theta) \) and \( \ell_i(\theta) \) is the log quasi conditional PDF of \( y_i | W_i \).

### 4 Simulation Experiments
In this section we conduct Monte Carlo experiments to investigate the finite sample performance of the proposed QML estimators in comparison with the OLS and GMM estimators. The bivariate system (1) is used as our data generating process (DGP), where the structural innovations are generated from the standardized \( \chi^2(10) \) variates such that the Gaussian likelihood for the QML is indeed misspecified. For our QML, while the log conditional variances can be any twice continuously differentiable function, we specify conditional variances as exponential and linear functions of exogenous variables (the two most commonly used functional forms for heteroskedasticity). They have a long history in econometrics, although they are typically seen in the context of solving the “problem” of heteroskedasticity rather than the context of assisting identification. For instance, see Glejser (1969), Goldfeld and Quandt (1965) and Rutemiller and Bowers (1968) for exponential variances, and Park (1966), Lancaster (1968) and Geary (1966) for the linear forms. Specifically, the data is simulated as follows:

i. Two independent \( \chi^2(10) \) random variables are generated and standardized to mean-zero and variance-one variables \( u_{kt} = [\chi^2(10) - 10]/\sqrt{20} \), for \( k = 1, 2 \) and \( i = 1, 2, ..., n \) with \( n = 500 \). Structural error terms \( \varepsilon_{kt} \) are then constructed as \( \varepsilon_{kt} = u_{kt}\sqrt{g_{kt}} \), where \( g_{kt} \)’s are the conditional variances described below.

ii. The vector of exogenous variables in the mean equations is set to \( x_i = [1, w_i]’ \), where \( w_i \sim \text{iid } N(0,1) \). The vector of exogenous variables in the conditional variances is set to \( z_i = x_i \).

iii. The vector of endogenous variables \( y_i = [y_{1i}, y_{2i}]’ \) is generated by the reduced-form system \( y_i = \Gamma^{-1}Bx_i + \Gamma^{-1}\varepsilon_i \), where \( \varepsilon_i = [\varepsilon_{1i}, \varepsilon_{2i}]’ \) is given by i. above. The matrices \( B \) and \( \Gamma \) are set to \( B = \begin{bmatrix} 0.5 & 0.7 \\ 0.8 & 0.6 \end{bmatrix} \) and \( \Gamma = \begin{bmatrix} 1 & 0.5 \\ -1 & 1 \end{bmatrix} \), i.e., \( y_{12} = -0.5 \) and \( y_{21} = 1 \) (see (6)).

iv. The conditional variances are generated by two alternative functional forms,

a) Linear form in squares: \( g_{1i} = 0.1 + 0.9w_i^2 \) and \( g_{2i} = 0.3 + 0.5w_i^2 \);

b) Exponential form: \( g_{1i} = \exp(0.1 + 0.9w_i) \) and \( g_{2i} = \exp(0.3 + 0.5w_i) \).

The DGP in a) will be called the “linear variance” and the DGP in b) the “exponential variance”.

With each of the two DGPs described above, 5,000 samples are generated. For each sample, which consists of 500 observations, we estimate the parameters using the OLS, the QML with
the log likelihood function (4), and the GMM of Lewbel (2012). The DGP parameter values are arbitrarily chosen, as commonly practised in the literature.

While the QML acknowledges the possibility of misspecification in the likelihood function (e.g. the normal likelihood function is used instead of the chi-squared likelihood that generates the data), in practice the functional forms of the conditional variances can be also misspecified. Thus, we are interested in the robustness of the QML to the misspecifications in the conditional variances. In particular, for each simulated sample, we estimate the parameters under five alternative specifications for the conditional variances

- QML1: linear in squares: \( g_{ki} = \alpha_{k0} + \alpha_{k1}w_i^2 \),
- QML2: exponential in levels: \( g_{ki} = \exp(\alpha_{k0} + \alpha_{k1}w_i) \),
- QML3: exponential in squares: \( g_{ki} = \exp(\alpha_{k0} + \alpha_{k1}w_i^2) \),
- QML4: exponential in levels and squares: \( g_{ki} = \exp(\alpha_{k0} + \alpha_{k1}w_i + \alpha_{k2}w_i^2) \),
- QML5: linear in levels and squares: \( g_{ki} = \alpha_{k0} + (\alpha_{k1}w_i + \alpha_{k2})^2 \),

We note that QML4 nests QML2 and QML3, while QML5 nests QML1. With possibly misspecified conditional variances, we are able to check the robustness of the QML estimators of the parameters in the mean equations that are correctly specified.

We use two sets of moments for the GMM estimation. Denote \( x_i = [1, w_i^2]' \), \( z_i = [w_i, w_i^2]' \), \( \varepsilon_{1i}(\theta) = y_{1i} - \gamma_{12}y_{2i} - \beta_{1i}^T x_i \), and \( \varepsilon_{2i}(\theta) = y_{2i} - \gamma_{21}y_{1i} - \beta_{2i}^T x_i \). The two sets of moments are defined as

- GMM1: \( q_{1i} = x_i \varepsilon_{1i}(\theta), q_{2i} = x_i \varepsilon_{2i}(\theta), q_{3i} = z_i \varepsilon_{1i}(\theta) \varepsilon_{2i}(\theta), q_{4i} = \varepsilon_{1i}(\theta) \varepsilon_{2i}(\theta) \); 
- GMM2: \( q_{1i} = x_i \varepsilon_{1i}(\theta), q_{2i} = x_i \varepsilon_{2i}(\theta), q_{3i} = (z_i - \pi) \varepsilon_{1i}(\theta) \varepsilon_{2i}(\theta), q_{4i} = (z_i - \pi) \varepsilon_{1i}(\theta) \varepsilon_{2i}(\theta) \); 

where \( \pi \) is the mean of \( z_i \). GMM2 includes exactly the same set of moments used in Lewbel (2012, p72). While GMM2 is designed for the normalisation rule \( \text{cov}(z_i, \varepsilon_{1i} \varepsilon_{2i}) = 0 \), it is also applicable to the normalisation used in this paper, i.e., \( \text{cov}(\varepsilon_{1i}, \varepsilon_{2i}) = 0 \). Being a variant of GMM2, GMM1 is tailored to the normalisation rule \( \text{cov}(\varepsilon_{1i}, \varepsilon_{2i}) = 0 \). In this setting, GMM1 contains 7 moments for 6 mean equation parameters, whereas there are 8 moments in GMM2 for 6 mean equation parameters and 2 extra parameters in \( \pi \), which are the means of \( w_i \) and \( w_i^2 \).

With the 5,000 replications, we calculate the biases and RMSEs for each estimator. The results are presented in Table 1 (for linear variance DGP) and Table 2 (for exponential variance DGP). As expected, OLS produces largest biases under both DGPs.
In Table 1, we observe that QML1, which correctly specify the conditional variances of the DGP, produce the best performance that is closely matched by a more general QML5. In estimating the mean equation parameters, the performances of QML1 and QML5 are followed by QML3 and QML4, which have substantially misspecified conditional variances. Importantly, the performances of QML3 and QML4 are markedly superior to that of GMM1. The conditional variances of QML2 are severely misspecified (using exponential functions of $w_i$ rather than linear functions of $w_i^2$), which lead to a performance slightly inferior to that of GMM1. GMM2 appears to be the worst performer in terms of RMSE in Table 1.

Turning to Table 2, the best performing estimators appear to be QML2 and QML4, which correctly specify the conditional variances. Importantly, although QML5 has substantially misspecified conditional variances, its performance in estimating mean equation parameters is as good as GMM1 that is correctly specified. QML3, with severely misspecified conditional variances (using $w_i^2$ rather than $w_i$), produces the mean equation parameter estimates with biases about halfway between those of GMM1 and GMM2. QML1, also severely misspecified (using linear function of $w_i^2$), and GMM2 yield largest biases and RMSE’s.

The simulation results appear to suggest that the QML estimators of the mean equation parameters perform well as long as the conditional variances are directionally correct. By “directionally correct”, for instance, we mean that, when the true conditional variances are monotone functions of $w_i^2$, the model conditional variances are also monotone functions of $w_i^2$. It appears that the correctness of the functional forms, linear or exponential, is not vital for the performance of the QML estimators of the mean equation parameters. These are evidently demonstrated by the robust performances of QML4 and QML5 under both DGPs in Tables 1 and 2. These results highlight the improvements gained by exploiting the knowledge about the conditional variances, even when they are mildly-misspecified.

The assumption A3 of Lewbel (2012, p71) requires $\text{cov}(x_i, \varepsilon_{ki}^2) \neq 0$ for at least one of $k = 1,2$. However, in the case of the linear variance DGP when $x_i \sim N(0,1)$ and $x_i$ is independent of $u_{ki} = \varepsilon_{ki}/g_{ki}^{1/2}$, we find $\text{cov}(x_i, \varepsilon_{ki}^2) = E[x_i(\varepsilon_{ki}^2 - E(\varepsilon_{ki}^2))] = E(x_i g_{ki} u_{ki}^2) = E(x_i[\alpha_k0 + \alpha_k1x_i^2]u_{ki}^2) = [\alpha_k0 E(x_i) + \alpha_k1 E(x_i^3)] E(u_{ki}^2) = 0$ for both $k = 1,2$. This is an example where the identifying conditions fail to hold for the GMM approach but continue to hold for the QML approach. This observation also explains the poor performance of GMM1 and GMM2 in Table 1.
In summary, the simulation results show that the QML estimators with correct conditional variances perform well in terms of biases and RMSEs. For the mean equation parameters, the QML estimators are robust to mild misspecifications in the conditional variances (e.g., QML4 under the linear variance DGP), although severe misspecifications (e.g., QML2 under the linear variance DGP) can substantially worsen the performance of the QML estimators. It appears that QML4 and QML5, which are “directionally correct”, are favoured when the true conditional variances are unknown.

The computation for the experiments of this section is carried out in Ox6.20. We use the MaxSQP function, which implements the Sequential Quadratic Programming algorithm, to numerically optimise the objective functions of both QML and GMM. As discussed under (2), there are two true parameter points for the SES. To focus on the parameter point in (1), for both QML and GMM, we restrict the sign of \( \gamma_{12} \) in the first equation to be negative such that it is always the “demand curve”. This is done by writing \( \gamma_{12} = -\gamma_0^2 \). Further, as the numerical optimisation may fail for some samples, we apply a filtering process for both QML and GMM as follows: (a) estimate the parameters; (b) if both estimates of the \( \gamma \) parameters are in the interval (-10, 10) then keep the estimates; (c) otherwise change the starting parameter values and go back to step (a) for up to three times. If after three times of altering the starting value and estimating the model, the magnitude of \( \gamma \) estimates is still in excess of 10, the last set of estimates is kept. A sample is discarded if MaxSQP algorithm fails to converge for one or more estimation methods.

5 Application: Stock Price Elasticities

A number of propositions in finance (including the Modigliani-Miller irrelevance theorems, CAPM and APT) rely on horizontal demand and supply curves for stocks, so that traded volumes do not affect prices. However, there is a growing body of research that documents empirical evidence for less than perfectly elastic demand and supply curves. For instance, Bagwell (1992) constructs upward-sloping supply curves using individual shareholder bids from 32 Dutch auction stock repurchases. Similar evidence is found for inter-firm tender offers (Bradley et al., 1988), fixed-price tender offers (Brown and Ryngaert, 1992), and large block transactions (Holthausen et al., 1990). On the other hand, evidence for downward-sloping demand curves is reported in Shleifer (1986), who studies the demand for stocks newly included in the S&P 500 index, while controlling for informational content such
inclusions may convey. Kaul et al. (2000) provides corroborative evidence of less than perfectly elastic demand from the Toronto Stock Exchange, while Petajisto (2011), and Wurgler and Zhuravskaya (2002) present more extensive results for the Russell 2000 and S&P 500 stocks. Kalay et al (2004) find elastic demand and supply schedules based on the pre-opening orders. Overall, this literature has so far chiefly relied on event studies to identify demand or supply schedules.

Differing from the existing studies, we treat traded prices and volumes of stocks as the endogenous variables in the SES given in (1), while controlling for stock-specific characteristics. The control variables include: book-to-market value of equity, market beta, and earnings before interest, tax, depreciation and amortization (EBITDA). These factors are motivated by the empirical literature that seeks to explain equity returns, see Basu (1977), Banz (1981), and Fama and French (1992) among others. Our first two factors, book-to-market equity and market beta, have been used extensively in the literature since they were introduced by Fama and French (1992). Our third factor, EBITDA, simultaneously accounts for two firm characteristics: company size and earnings potential.

We write our SES model as follows

\[ p_t = \beta_{10} + \gamma_1 q_t + \beta_1 x_t + \varepsilon_{1t}, \]

\[ q_t = \beta_{20} + \gamma_2 p_t + \beta_2 x_t + \varepsilon_{2t}, \]

where \( p_t \) and \( q_t \) are the natural logarithms of the closing prices and traded volumes on a particular day; \( x_t \) is a \((3 \times 1)\) vector comprising EBITDA, market beta, and book-market equity. The variables in \( x_t \) are standardized \(^5\) to have zero mean and unit variance. Given the remarks surrounding equation (2), a priori information is needed to label the supply and demand curves. Following the literature on stock price elasticities (Bagwell (1992) among others), we label the supply schedule as the equation associated with a positive \( \gamma \), and the demand schedule as the other equation. As \( \varepsilon_{1t} \) and \( \varepsilon_{2t} \) are normalised to be uncorrelated (Assumption A3), \( \varepsilon_{1t} \) is interpreted as the shock that shifts the supply curve (suppose \( \gamma_1 > 0 \)) without altering the demand. Hence, \( \gamma_2 \) and \( 1/\gamma_1 \) are the price elasticities of demand and supply. The hypothesis of interest is \( H_0: \gamma_2 = 1/\gamma_1 = 0 \), i.e., the price is not affected by the volume.

\(^5\)The standardisation involves subtracting the sample mean and dividing by the sample standard deviation. It is done for two reasons: i) there are significant differences in the distributional characteristics of the three variables (see Table 3 for their summary statistics), and ii) these variables are exponentiated in the variance equations (11) below and, if not standardised, can lead to overly large values that cause difficulties in the maximization of the likelihood function.
In this context, because it is implausible to argue \textit{a priori} that one or more variables in $x_i$ should be excluded from one equation or the other in (10) for identification purposes, we need to exploit the heteroskedasticity in the error terms to locally identify the parameters in (10). Given that QML4 is shown to have relatively robust performances in Section 4, we specify the functional forms of the heteroskedasticity as

$$g_{ki} = \text{var}(\varepsilon_{ki}|W_i) = \exp(\alpha_{k0} + \alpha'_{k1}x_i + \alpha'_{k2}x_i^2)$$ \quad \text{for} \quad k = 1, 2,$$

where $x_i$ is defined as in (10), $x_i^2$ is the vector containing squared elements of $x_i$ and $(\alpha_{k0}, \alpha_{k1}, \alpha_{k2})$ are parameters with conforming dimensions.

Our dataset consists of 1,902 companies included in the Russell 3000 index for which we were able to obtain all necessary data from Bloomberg. Traded prices and volumes are obtained for an arbitrarily selected day – August 30, 2012 – as well as for the days before and after (August 29 and August 31). Of the explanatory variables, EBITDA is the June 2012 quarter figure and is exogenous or predetermined with respect to the endogenous variables $p_i$ and $q_i$. Market betas are computed by Bloomberg using two years of daily data ending on August 29, 2012, while book-market equity is computed using June 2012 quarter book values, and August 29, 2012 prices. Thus, these two factors are also predetermined with respect to the endogenous prices and volumes on August 30, 2012. The summary statistics are given in Table 3.

Table 4 presents parameter estimates corresponding to the market conditions on August 30, 2012. As previously discussed, we label the equation with a positive $\gamma$ estimate as the supply schedule, which is reported in the first column of the table with $\hat{\gamma}_1 = 0.426$. Given that the robust standard error for $\hat{\gamma}_1$ is 0.086, the hypothesis $H_0: \gamma_1 = \gamma_2^{-1} = 0$ is strongly rejected. The point estimate of the supply price elasticity is 2.347 with the standard error 0.474 (by the delta method). The second column of Table 4 contains the estimates of the demand equation associated with the negative structural coefficient $\hat{\gamma}_2 = -1.149$ on the log price with the standard error 0.220, which measures the demand price elasticity and is highly significant. We observe that $\hat{\gamma}_2^{-1} = 0.870$ with the standard error 0.167 also strongly rejects $H_0$. The supply and demand curves are graphically presented in Figure 1, where two more sets of curves associated with market conditions on the two days surrounding August 30, 2012 are also presented. These curves are clearly not horizontal. Shifts across the three sets of curves illustrate adjustments to news arrivals over the three adjacent days.
Turning to the variance equations, it is evident that most of the coefficients in (11) are statistically significant with different magnitudes. In particular, the coefficients on EBITDA and EBITDA$^2$ have opposite signs in the two conditional variances. This confirms that the local identifying condition is met as the $g_{2i}$ is not proportional to $g_{2i}$. As the coefficients on EBITDA and EBITDA$^2$ are respectively negative and positive in the supply-shock conditional variance, the variation in the supply shock is positively (negatively) related to EBITDA for stocks with standardised EBITDA greater (less) than 0.202/2, ceteris paribus.

The third panel in Table 4 gives the Breusch-Pagan heteroskedasticity test statistics for the reduced-form errors and the standardised structural errors (i.e., $\varepsilon_{kl}/g_{kl}^{1/2}$) respectively. Under the null hypothesis of homoskedasticity, the Breusch-Pagan test statistic is asymptotically distributed as $\chi^2(3)$ with the 5% critical value being 7.815. Clearly, based on the test statistics in Table 4, there is strong evidence for heteroskedasticity in the reduced-form errors and there is none in the standardised structural errors. These test statistics suggest that the heteroskedasticity in the structural errors is well captured by the conditional variances used in our estimation. This observation suggests that the possible misspecification in the conditional variances is at most mild.

To check the robustness of the results in Table 4, we also estimate the SES with the extended linear specification (QML5) for the conditional variances

$$g_{ki} = \alpha_{k0} + (\alpha'_{k1}x_i + \alpha_{k2}^2)^2 \quad \text{for} \quad k = 1,2,$$

where $x_i$ is defined in (10). The results are reported in Table 5. The mean equation estimates in Table 5 are qualitatively the same as those in Table 4. While the Breusch-Pagan test statistics on the standardised structural error terms are substantially improved over those on the reduced-form error terms, they still indicate that some heteroskedasticity remains in the demand shock (a reason we prefer the results in Table 4).

Furthermore, with the benchmark specification (11), Table 6 presents the estimation results with logarithms of EBITDA and Book-Market (without the standardisation documented in footnote 6). The mean equation estimates in Table 6 are similar to those in Table 4 in terms of signs. They confirm that the hypothesis $H_0: \gamma_1 = \gamma_2^{-1} = 0$ is also strongly rejected. However, the variance equation estimates differ from those in Table 4, prominently in the demand-shock conditional variance. As Tables 4 and 6 are based on two types of data transformations, the difference in the variance parameter estimates signifies that the demand-shock conditional variance behaves differently under the two data transformations.
In summary, using a SES to model the traded prices and volumes in the Russell 3000 index, we find that the point estimates of the supply and demand elasticities are finite and nonzero with small standard errors. This provides new evidence against the hypothesis that traded volume does not affect the price of a stock. We acknowledge that there may be other relevant factors, such as the momentum (e.g. Carhart, 1997) and liquidity (e.g. Amihud and Mendelson 1986; Brennan and Subrahmanyam 1996) that need to be accounted for, although we limit our analysis to the three factors in this paper owing to the limitations of our data sources and leave the extensions for future research.

6 Conclusions

We examine the identification and estimation of simultaneous equation systems (SES) with heteroskedasticity in the framework of the Gaussian quasi maximum likelihood (QML) method. Our results are general and applicable to any SES where the conditional variances can be parameterised as twice continuously differentiable functions. We provide a set of sufficient conditions for the local identification of the structural parameters and derive the consistency and asymptotic normality of the QML estimators.

Simulation experiments indicate that the QML performs well in finite samples. Even the QML with mildly-specified conditional variances delivers performances comparable to those of correctly-specified GMM. We observe that increasing the accuracy of the conditional variance specifications leads to improvements in the precision of the QML estimators of the mean equation parameters.

Our empirical application is the first study that directly estimates demand and supply elasticities for listed equities. By modelling traded prices and volumes on a cross-section of stocks included in the Russell 3000 index, we find strong evidence against the hypothesis of homogenous valuations (i.e. the hypothesis that equities have perfectly elastic demand and supply schedules). Our results support the conclusion of a number of existing articles that rely on event studies, an empirical method vastly different from ours.

Directions for future research may include investigating the possibility of using semiparametric conditional heteroskedasticity in the context of QML. Nevertheless, given the difficulties in implementing the semiparametric approach to modelling the conditional variances (see Farre et al (2013)), a systematic examination of this issue is beyond the scope of the current paper.
7 Appendix

Proof of Theorem 1. The claim of Theorem 1 is implied by Theorem 2, Example 1 in Section 2.2 and Lemma 1 below.

Proof of Theorem 2. The proof involves (i) finding the Hessian matrix of the log quasi likelihood; (ii) deriving the expected value of the Hessian and expressing it in the form of (8).

It is easy to verify the following differentiations
\[
\frac{d\ln{|\Gamma|}}{d\alpha} = \text{tr}(d\Gamma^{-1}),
\]
\[
\frac{d\ln{|G_i|}}{d\alpha} = d\sum_{k=1}^{p} F_k(z_i, \alpha_k) = [f'_{11}, ..., f'_{pi}]d\alpha,
\]
\[
\frac{d}{d\alpha}[G_i^{-1}e_i(\theta)e_i'(\theta)] = 2\text{tr}[d\epsilon_i(\theta)e_i'(\theta)G_i^{-1}] - \left[f'_{11} \frac{\epsilon^2_1(\theta)}{g_{11}}, ..., f'_{pi} \frac{\epsilon^2_p(\theta)}{g_{pi}}\right]d\alpha,
\]
where \(\alpha = [\alpha'_1, ..., \alpha'_p]'\). These lead to the differentiation of (7),
\[
dL_n = \text{tr}(d\Gamma[n\Gamma^{-1} - \Sigma_{i=1}^{n} y_i e_i(\theta)G_i^{-1}]) + \text{tr}(dB \Sigma_{i=1}^{n} x_i e_i(\theta)G_i^{-1})
\]
\[
\quad + \frac{1}{2} \Sigma_{i=1}^{n}[f'_{11} \left(\frac{\epsilon^2_1(\theta)}{g_{11}} - 1\right), ..., f'_{pi} \left(\frac{\epsilon^2_p(\theta)}{g_{pi}} - 1\right)]d\alpha,
\]
which, together with the formula \(\text{tr}(XY) = \text{vec}(Y)'\text{vec}(X')\), leads to the gradients
\[
\nabla_F = \frac{\partial dL_n}{\partial (y_{12}, ..., y_{1p}, ..., y_{p,p-1})} = R_p'\text{vec}(\Sigma_{i=1}^{n} y_i e_i(\theta)G_i^{-1} - n\Gamma^{-1}),
\]
\[
\nabla_B = \frac{\partial dL_n}{\partial \text{vec}(\Gamma')} = \text{vec}(\Sigma_{i=1}^{n} x_i e_i(\theta)G_i^{-1}),
\]
\[
\nabla_\alpha = \frac{\partial dL_n}{\partial \alpha} = \frac{1}{2} \Sigma_{i=1}^{n}[f'_{11} \left(\frac{\epsilon^2_1(\theta)}{g_{11}} - 1\right), ..., f'_{pi} \left(\frac{\epsilon^2_p(\theta)}{g_{pi}} - 1\right)]',
\]
where \((y_{12}, ..., y_{1p}, ..., y_{p,1}, ..., y_{p,p-1})\) are free parameters in \(\Gamma\) (row-wise or equation-wise),
\[
\frac{\partial \text{vec}(\Gamma')}{\partial (y_{12}, ..., y_{1p}, ..., y_{p,1}, ..., y_{p,p-1})} = -R_p = -\begin{bmatrix} I_{p-1} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & I_{p-k}\end{bmatrix}
\]
and \(I_{p-k}\) is the \(p\)-dimensional identity matrix \(I_p\) with the \(k\)th column being deleted. Let \(K_p\) be the invertible \(p^2 \times p^2\) matrix such that \(\text{vec}(X) = K_p \text{vec}(X')\) for any \(p \times p\) matrix \(X\). In fact, \(K_p = [I_p \otimes e_1, ..., I_p \otimes e_p]\) that satisfies \(K'_p = K_p\), \(K^{-1}_p = K_p\) and \((X \otimes Y)K_p = K_p(Y \otimes X)\) for any \(p \times p\) matrices \(X\) and \(Y\), where \(e_k\) is the \(k\)th column of \(I_p\). Notice also
Differentiating the above gradients gives the sub-matrices of the Hessian at the true parameter point \( \theta_0 \)

\[
\nabla_{tt} = \frac{\partial v_t}{\partial \gamma_1} = -R_p \left( \sum_{i=1}^{n} \left( G_{1i}^{-1} \otimes y_i y_i' \right) + n (\Gamma_1^{-1} \otimes \Gamma_1^{-1} K_p) \right) R_p,
\]

\[
\nabla_{tt} = \frac{\partial v_t}{\partial \gamma_i} = -R_p \left( \sum_{i=1}^{n} \left( G_{1i}^{-1} \otimes y_i x_i' \right) \right),
\]

\[
\nabla_\alpha = \frac{\partial v_\alpha}{\partial \gamma_i} = -R_p \left( \sum_{i=1}^{n} \left( \text{diag}(\gamma_i) G_{1i}^{-1} \phi_i' \otimes y_i \right) \right) = -R_p \left( \sum_{i=1}^{n} \left( \text{diag}(\gamma_i) G_{1i}^{-1} \otimes y_i \right) \phi_i' \right),
\]

\[
\nabla_{BB} = \frac{\partial v_B}{\partial \gamma_i} = -\sum_{i=1}^{n} \left( G_{1i}^{-1} \otimes x_i x_i' \right),
\]

\[
\nabla_B = \frac{\partial v_B}{\partial \gamma_i} = -\sum_{i=1}^{n} \left( \text{diag}(\gamma_i) G_{1i}^{-1} \phi_i' \otimes x_i \right) = -\sum_{i=1}^{n} \left( \text{diag}(\gamma_i) G_{1i}^{-1} \otimes x_i \right) \phi_i',
\]

\[
\nabla_{aa} = \frac{\partial v_a}{\partial \gamma_i} = -\frac{1}{n} \sum_{i=1}^{n} \text{diag} \left( f_i \frac{\epsilon_{1i}}{\gamma_{1i}} f_i', \ldots, f_p \frac{\epsilon_{pi}}{\gamma_{pi}} f_p' \right),
\]

\[
\quad + \frac{1}{n} \sum_{i=1}^{n} \text{diag} \left( \frac{\partial f_i}{\partial \gamma_{1i}} \left( \frac{\epsilon_{1i}^2}{\gamma_{1i}} - 1 \right), \ldots, \frac{\partial f_p}{\partial \gamma_{pi}} \left( \frac{\epsilon_{pi}^2}{\gamma_{pi}} - 1 \right) \right),
\]

where \( \phi_i' = \text{diag} \left( f_i', \ldots, f_p' \right) \) and \( \epsilon_i' dG_i^{-1} = -d\alpha_i' \phi_i G_i^{-1} \text{diag}(\gamma_i) \). Let \( m_k \) be the \( k \)th column of \( \Gamma_1^{-1} \) for \( k = 1, \ldots, p \). It can be easily verified that at \( \theta_0 \)

\[
E(\text{diag}(\gamma_i) G_i^{-1} \otimes y_i | W_i) = E(\text{diag}(\gamma_i) \otimes \Gamma_1^{-1} \varepsilon_i | G_i^{-1} | W_i) = \text{diag}(\bar{m}_1, \ldots, \bar{m}_p),
\]

\[
E(\text{diag}(\gamma_i) G_i^{-1} \otimes x_i | W_i) = 0,
\]

\[
E(G_i^{-1} \otimes y_i x_i' | W_i) = (G_i^{-1} \otimes \Gamma_1^{-1} B x_i x_i'),
\]

\[
E(G_i^{-1} \otimes y_i y_i' | W_i) = (G_i^{-1} \otimes \Gamma_1^{-1} B x_i x_i' B' \Gamma_1^{-1}) + (G_i^{-1} \otimes \Gamma_1^{-1} G_i \Gamma_1^{-1}),
\]

\[
E(f_i \frac{\epsilon_{1i}}{\gamma_{1i}} f_i' | W_i) = f_k f_k', \quad E \left[ \frac{\partial f_i}{\partial \gamma_{1i}} \left( \frac{\epsilon_{1i}^2}{\gamma_{1i}} - 1 \right) | W_i \right] = 0, \quad \text{for} \quad k = 1, \ldots, p.
\]

By iterated expectations, the quasi Fisher information matrix, \( J_{\theta \theta} = -E(\nabla_{\theta \theta}/n) \), at \( \theta_0 \) consists of the following sub-matrices

\[
J_{tt} = R_p \left\{ \frac{1}{n} \sum_{i=1}^{n} (G_i^{-1} \otimes y_i y_i') + (\Gamma_1^{-1} \otimes \Gamma_1^{-1}) K_p \right\} R_p
\]

\[
= R_p \left\{ \frac{1}{n} \sum_{i=1}^{n} \left( (G_i^{-1} \otimes \Gamma_1^{-1} B x_i x_i' B' \Gamma_1^{-1}) + (G_i^{-1} \otimes \Gamma_1^{-1} G_i \Gamma_1^{-1}) \right) \right\}
\]

\[
\quad + (\Gamma_1^{-1} \otimes \Gamma_1^{-1}) K_p \right) R_p
\]

\[
= R_p \left( I_p \otimes \Gamma_1^{-1} \right) \left\{ \frac{1}{n} \sum_{i=1}^{n} \left( (I_p \otimes B)(G_i^{-1} \otimes x_i x_i')(I_p \otimes B') + (G_i^{-1} \otimes G_i) \right) \right\}
\]

\[
\quad + K_p \right) (I_p \otimes \Gamma_1^{-1}) R_p
\]

\[
J_{GB} = R_p \frac{1}{n} \sum_{i=1}^{n} (G_i^{-1} \otimes \Gamma_1^{-1} B x_i x_i') = R_p \left( I_p \otimes \Gamma_1^{-1} \right) \left( I_p \otimes B \right) H_x.
\]

24
It follows from elementary operations that $J_{\theta \theta}$ is positive definite if and only if $J_{\Gamma \cdot \alpha}$ is, where

$$J_{\Gamma \cdot \alpha} = J_{\Gamma \Gamma} - J_{\Gamma B} J_{B B}^{-1} J_{B \Gamma} = J_{\Gamma B} J_{\alpha \alpha}^{-1} J_{\alpha \Gamma}$$

$$J_{\Gamma \cdot \alpha} = R'_p (I_p \otimes \Gamma^{-1}) [E \sum_{i=1}^{\infty} G_i^{-1} \otimes G_i + K_p] (I_p \otimes \Gamma^{-1}^i) R_p$$

$$= 2R'_p \diag[\bar{m}_1, ..., \bar{m}_p] \phi^2 H_f^{-1} \phi \diag[\bar{m}_1', ..., \bar{m}_p'] R_p.$$
We complete the proof by showing that $\psi = V\Psi V'$ is diagonal and positive semi-definite. Firstly, the upper diagonal blocks of $V\Psi$, the $(k, l)$th blocks for $k < l$, are all zero,

$$\Psi_{kl} + v_{kl}A_l + \sum_{j=k+1}^p v_{kj}\Psi_{jl} = 0,$$

because $e'_{k-j}e_{l-j} = 0$ for all $j > k$ and $l > k$. The diagonal blocks of $V\Psi$ are given by

$$\psi_k = A_k - \sum_{l=k+1}^p \psi_{kl}A_l^{-1}\psi_{lk} = A_k - \sum_{l=k+1}^p e_{l-k}e'_{l-k}\lambda_{k,l}^{-1} = \text{diag}[\lambda_{1,k}, ..., \lambda_{k-1,k}, (\lambda_{k+1,k} - \lambda_{k,k+1}^{-1}), ..., (\lambda_{p,k} - \lambda_{k,p}^{-1})], \quad k = 1, ..., p.$$

The lower diagonal blocks of $V\Psi$, the $(k, l)$th blocks for $k > l$, are those of $\Psi$,

$$\Psi_{kl} + \sum_{j=k+1}^p v_{kj}\Psi_{jl} = \Psi_{kl},$$

because $e'_{k-j}e_{l-j} = 0$ for all $j > k$ and $l < k$. Secondly, the lower diagonal blocks of $\psi = V\Psi V'$, the $(k, l)$th blocks for $k < l$, are all zero,

$$\psi_{kl} + \sum_{j=k+1}^{l-1} \psi_{kj}v_{lj} + \psi_{jk}v'_{lk} = \psi_{kl} + \psi_{jk}v'_{lk} = \psi_{kl} - \psi_{jk}\psi_{kl}^{-1} = 0.$$

The diagonal blocks of $\psi$ are obviously those of $V\Psi$ and can be detailed as

$$\psi_p = A_p,$$

$$\psi_{p-1} = A_{p-1} - e_{p-1}e'_{p-1}\lambda_{p-1,p}^{-1} = \text{diag}[\lambda_{1,p-1}, ..., \lambda_{p-2,p-1}, (\lambda_{p-1,p} - \lambda_{p-1,p}^{-1})],$$

$$\vdots$$

$$\psi_2 = A_2 - \sum_{j=3}^p e_{j-2}e'_{j-2}\lambda_{j,j}^{-1} = \text{diag}[\lambda_{1,2}, (\lambda_{3,2} - \lambda_{2,3}^{-1}), ..., (\lambda_{p-2,j} - \lambda_{p-1,p}^{-1})],$$

$$\psi_1 = A_1 - \sum_{j=2}^p e_{j-1}e'_{j-1}\lambda_{1,j}^{-1} = \text{diag}[\lambda_{21}, ..., (\lambda_{p-1,1} - \lambda_{1,p}^{-1})].$$

Finally, by Lemma 1, $\lambda_{kl} - \lambda_{lk}^{-1} \geq 0$ confirms that $\psi$ is positive semi-definite. $\blacksquare$

**Proof of Theorem 3.** The proof is mainly about verifying the conditions required for the theorems in Newey and McFadden (1994) and the uniform WLLN in Bierens (2004). Note that $L_n(\theta) = \sum_{i=1}^n \ell_i(\theta)$ with

$$\ell_i(\theta) = \frac{p}{2} \ln(2\pi) + \ln(|\Gamma|) - \frac{1}{2} \ln(\{|G_i|\} + \text{tr}[G_i^{-1}\epsilon_i(\theta)\epsilon_i(\theta)']),$$

where $\epsilon_i(\theta) = \Gamma y_i - Bx_i = \Gamma [G_0^{-1}(B_0x_i + \epsilon_i)] - Bx_i$. By C3 and the compactness of $\Theta_\delta$, $\frac{1}{n}L_n(\theta)$ converges in probability to $\ell_0(\theta) = E\ell_i(\theta)$ uniformly in $\Theta_\delta$. Also by C3 and compactness of $\Theta_\delta$, $\ell_i(\theta)$, $\frac{\partial \ell_i(\theta)}{\partial \theta}$ and $\frac{\partial^2 \ell_i(\theta)}{\partial \theta \partial \theta'}$ are all continuous and bounded in $\Theta_\delta$. It follows that
E(∂ \frac{1}{n} L_n(\theta)/\partial \theta) = \partial \ell_0(\theta)/\partial \theta \quad \text{and} \quad E(\partial^2 \frac{1}{n} L_n(\theta)/\partial \theta \partial \theta') = \partial^2 \ell_0(\theta)/\partial \theta \partial \theta'.

As \(-E(\partial^2 \frac{1}{n} L_n(\theta)/\partial \theta \partial \theta')\) is positive definite at \(\theta_0\) under the conditions of Theorem 2, \(\ell_0(\theta)\) is concave in \(\Theta_\delta\). Consequently, \(\theta_0\) is the unique maximiser of \(\ell_0(\theta)\) in \(\Theta_\delta\) because the concavity of \(\ell_0(\theta)\) and \(E(\partial \ell_i(\theta_0)/\partial \theta) = 0\), which is verified in Lemma 2. The consistency, \(\hat{\theta}_n \xrightarrow{p} \theta_0\), is then established from Theorem 2.1 of Newey and McFadden (1994). For the asymptotic normality of \(\hat{\theta}_n\), we note that \(\partial \ell_i(\theta_0)/\partial \theta\) satisfies the CLT by Lemma 2. By C3 and the uniform WLLN, \(\partial^2 \frac{1}{n} L_n(\theta)/\partial \theta \partial \theta'\) converges in probability to its mean uniformly in \(\Theta_\delta\) because it is continuous and bounded in \(\Theta_\delta\). Finally, as \(J_{\theta\theta}\) is of full rank under the conditions of Theorem 2, the asymptotical normality of \(\hat{\theta}_n\) follows from Theorem 3.1 of Newey and McFadden (1994).

**Lemma 1 (Cauchy-Schwartz inequality).** Let \((u_i, v_i)\) be a pair of positive random variables for \(i = 1, \ldots, n\). Then \(E(\sum_{i=1}^{n} u_i / n)E(\sum_{i=1}^{n} v_i / n) \geq \left[E(\sum_{i=1}^{n} \sqrt{u_i v_i} / n)\right]^2\), where the equality holds if and only if \(v_i = c^2 u_i\) for a constant \(c\) and all \(i\).

**Proof of Lemma 1.** Let \(E_n(\cdot) = \sum_{i=1}^{n}(\cdot) / n\). Because

\[
E[E_n((\sqrt{v_i} - c\sqrt{u_i})^2)] = E[E_n(u_i)]c^2 - 2E[E_n(\sqrt{u_i v_i})]c + E[E_n(v_i)] \geq 0
\]

for any real \(c\), the coefficient of the quadratic form (in \(c\)) must satisfy the claimed inequality, where the equality holds if and only if \(\sqrt{v_i} = c\sqrt{u_i}\) for a constant \(c\) and all \(i\).

**Lemma 2.** Under Assumptions B and C, \(E[\partial \ell_i(\theta_0)/\partial \theta|W_i] = 0\) and the CLT applies to \(\partial \ell_i(\theta_0)/\partial \theta\).

**Proof of Lemma 2.** The gradient of \(\ell_i(\theta)\) at \(\theta_0\) can be expressed as follows

\[
\frac{\partial \ell_i(\theta_0)}{\partial (y_1, \ldots, y_{i-1}, x_{i-1} G_{i-1}^{-1}, \ldots, y_{p-1})} = R_p^{'} \text{vec}(y_i \epsilon_i G_{i0}^{-1} - \Gamma_0^{-1})
\]

\[
= R_p^{'} \text{vec}(\Gamma_0^{-1} B_0 x_i \epsilon_i G_{i0}^{-1} + \Gamma_0^{-1}(\epsilon_i \epsilon_i G_{i0}^{-1} - I_p))
\]

\[
= R_p^{'} (I_p \otimes \Gamma_0^{-1} B_0) \text{vec}(x_i \epsilon_i G_{i0}^{-1}) + (I_p \otimes \Gamma_0^{-1}) \text{vec}(\epsilon_i \epsilon_i G_{i0}^{-1} - I_p),
\]

\[
\frac{\partial \ell_i(\theta_0)}{\partial \text{vec}(\theta')} = \text{vec}(x_i \epsilon_i G_{i0}^{-1}),
\]

\[
\frac{\partial \ell_i(\theta_0)}{\partial \alpha} = \frac{1}{2} \left[f_{1i0} \left(\frac{\epsilon_i^2}{g_{1i0}} - 1\right), \ldots, f_{pi0} \left(\frac{\epsilon_i^2}{g_{pi0}} - 1\right)\right]'.
\]
where \((y_{12}, \ldots, y_{1p}, \ldots, y_{p1}, \ldots, y_{p,p-1})\) are free parameters in \(\Gamma\) (row-wise or equation-wise) and the quantities with 0 at the right-most subscript are those evaluated at \(\theta_0\). The claim of Lemma 2 follows from these expressions immediately. \(\blacksquare\)
8 References


### Table 1: Monte Carlo Simulation – Linear Variance DGP

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>True Parameter</th>
<th>OLS</th>
<th>GMM1</th>
<th>GMM2</th>
<th>QML1*</th>
<th>QML2</th>
<th>QML3</th>
<th>QML4</th>
<th>QML5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{12}$</td>
<td>-0.5</td>
<td>Bias</td>
<td>-0.832</td>
<td>-0.190</td>
<td>-0.160</td>
<td>0.015</td>
<td>-0.186</td>
<td>0.018</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.834</td>
<td>0.396</td>
<td>1.103</td>
<td>0.131</td>
<td>0.487</td>
<td>0.213</td>
<td>0.230</td>
</tr>
<tr>
<td>$\gamma_{21}$</td>
<td>1</td>
<td>Bias</td>
<td>0.504</td>
<td>0.226</td>
<td>0.191</td>
<td>-0.010</td>
<td>0.234</td>
<td>-0.002</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.511</td>
<td>0.390</td>
<td>1.906</td>
<td>0.150</td>
<td>0.420</td>
<td>0.217</td>
<td>0.232</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.5</td>
<td>Bias</td>
<td>-0.666</td>
<td>-0.152</td>
<td>-0.127</td>
<td>0.012</td>
<td>-0.148</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>-0.668</td>
<td>0.321</td>
<td>0.897</td>
<td>0.109</td>
<td>0.394</td>
<td>0.174</td>
<td>0.187</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.8</td>
<td>Bias</td>
<td>-0.777</td>
<td>-0.178</td>
<td>-0.147</td>
<td>0.014</td>
<td>-0.175</td>
<td>0.017</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.781</td>
<td>0.379</td>
<td>1.187</td>
<td>0.134</td>
<td>0.460</td>
<td>0.209</td>
<td>0.223</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>0.7</td>
<td>Bias</td>
<td>0.051</td>
<td>0.021</td>
<td>0.019</td>
<td>-0.001</td>
<td>0.026</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.064</td>
<td>0.054</td>
<td>0.144</td>
<td>0.036</td>
<td>0.058</td>
<td>0.040</td>
<td>0.040</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>0.6</td>
<td>Bias</td>
<td>0.168</td>
<td>0.076</td>
<td>0.063</td>
<td>-0.003</td>
<td>0.079</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.179</td>
<td>0.142</td>
<td>0.589</td>
<td>0.070</td>
<td>0.152</td>
<td>0.088</td>
<td>0.093</td>
</tr>
<tr>
<td>$\alpha_{10}$</td>
<td>0.1</td>
<td>Bias</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.001</td>
<td>-0.378</td>
<td>-1.385</td>
<td>-1.355</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.030</td>
<td>0.672</td>
<td>1.386</td>
<td>1.390</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>0.9</td>
<td>Bias</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.036</td>
<td>-0.903</td>
<td>-0.116</td>
<td>-0.902</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.202</td>
<td>0.910</td>
<td>0.149</td>
<td>0.907</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0</td>
<td>Bias</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.782</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.782</td>
<td>0.006</td>
</tr>
<tr>
<td>$\alpha_{20}$</td>
<td>0.3</td>
<td>Bias</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.006</td>
<td>-0.616</td>
<td>-1.199</td>
<td>-1.200</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.053</td>
<td>0.653</td>
<td>1.212</td>
<td>1.214</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>0.5</td>
<td>Bias</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.000</td>
<td>-0.497</td>
<td>-0.042</td>
<td>-0.497</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.106</td>
<td>0.509</td>
<td>0.082</td>
<td>0.505</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>0</td>
<td>Bias</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.457</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.462</td>
<td>0.060</td>
</tr>
<tr>
<td>$\pi_w$</td>
<td>0</td>
<td>Bias</td>
<td>-</td>
<td>-</td>
<td>-0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>-</td>
<td>-</td>
<td>0.048</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\pi_{w^2}$</td>
<td>1</td>
<td>Bias</td>
<td>-</td>
<td>-</td>
<td>0.003</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>-</td>
<td>-</td>
<td>0.106</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Notes:** * denotes a QML estimator with the correctly specified variance equation. The reported statistics are calculated from 5,000 replications of the experiment, using samples of 500 observations and the following formula: $Bias = \frac{1}{5000} \sum_{i=1}^{5000} (\hat{\theta}_i - \theta)$, $RMSE = \sqrt{\frac{1}{5000} \sum_{i=1}^{5000} (\hat{\theta}_i - \theta)^2}$. 
<table>
<thead>
<tr>
<th>( \theta )</th>
<th>True Parameter</th>
<th>OLS</th>
<th>GMM1</th>
<th>GMM2</th>
<th>QML1</th>
<th>QML2*</th>
<th>QML3</th>
<th>QML4</th>
<th>QML5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_{12} )</td>
<td>-0.5</td>
<td>Bias: -0.779</td>
<td>-0.053</td>
<td>-0.101</td>
<td>0.168</td>
<td>0.004</td>
<td>-0.074</td>
<td>0.003</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE: 0.781</td>
<td>0.212</td>
<td>0.661</td>
<td>0.862</td>
<td>0.170</td>
<td>0.360</td>
<td>0.175</td>
<td>0.272</td>
</tr>
<tr>
<td>( \gamma_{21} )</td>
<td>1</td>
<td>Bias: 0.565</td>
<td>0.074</td>
<td>0.123</td>
<td>-0.055</td>
<td>0.004</td>
<td>-0.120</td>
<td>0.006</td>
<td>-0.020</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE: 0.571</td>
<td>0.262</td>
<td>0.452</td>
<td>0.493</td>
<td>0.208</td>
<td>0.380</td>
<td>0.213</td>
<td>0.257</td>
</tr>
<tr>
<td>( \beta_{11} )</td>
<td>0.5</td>
<td>Bias: -0.624</td>
<td>-0.043</td>
<td>-0.080</td>
<td>0.133</td>
<td>0.002</td>
<td>0.060</td>
<td>0.001</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE: -0.625</td>
<td>0.181</td>
<td>0.539</td>
<td>0.690</td>
<td>0.145</td>
<td>0.294</td>
<td>0.148</td>
<td>0.226</td>
</tr>
<tr>
<td>( \beta_{12} )</td>
<td>0.8</td>
<td>Bias: -0.727</td>
<td>-0.049</td>
<td>-0.091</td>
<td>0.158</td>
<td>0.003</td>
<td>0.068</td>
<td>0.002</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE: 0.729</td>
<td>0.220</td>
<td>0.629</td>
<td>0.804</td>
<td>0.163</td>
<td>0.347</td>
<td>0.168</td>
<td>0.255</td>
</tr>
<tr>
<td>( \beta_{21} )</td>
<td>0.7</td>
<td>Bias: 0.058</td>
<td>0.008</td>
<td>0.014</td>
<td>-0.005</td>
<td>0.002</td>
<td>0.011</td>
<td>0.000</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE: 0.077</td>
<td>0.061</td>
<td>0.089</td>
<td>0.074</td>
<td>0.059</td>
<td>0.067</td>
<td>0.059</td>
<td>0.060</td>
</tr>
<tr>
<td>( \beta_{22} )</td>
<td>0.6</td>
<td>Bias: 0.189</td>
<td>0.025</td>
<td>0.042</td>
<td>-0.018</td>
<td>0.001</td>
<td>0.037</td>
<td>-0.001</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE: 0.200</td>
<td>0.109</td>
<td>0.170</td>
<td>0.169</td>
<td>0.084</td>
<td>0.141</td>
<td>0.086</td>
<td>0.095</td>
</tr>
<tr>
<td>( \alpha_{10} )</td>
<td>0.1</td>
<td>Bias: - - -</td>
<td>1.967</td>
<td>-0.005</td>
<td>-0.041</td>
<td>0.001</td>
<td>0.076</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE: - - -</td>
<td>6.197</td>
<td>0.233</td>
<td>0.430</td>
<td>0.244</td>
<td>0.188</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{11} )</td>
<td>0.9</td>
<td>Bias: - - -</td>
<td>0.062</td>
<td>0.012</td>
<td>-0.641</td>
<td>0.012</td>
<td>-0.581</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE: - - -</td>
<td>2.233</td>
<td>0.083</td>
<td>0.647</td>
<td>0.083</td>
<td>0.650</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{12} )</td>
<td>0</td>
<td>Bias: - - -</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.009</td>
<td>1.201</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE: - - -</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.059</td>
<td>2.110</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{20} )</td>
<td>0.3</td>
<td>Bias: - - -</td>
<td>1.267</td>
<td>-0.003</td>
<td>-0.022</td>
<td>-0.006</td>
<td>-0.038</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE: - - -</td>
<td>1.518</td>
<td>0.152</td>
<td>0.239</td>
<td>0.165</td>
<td>1.264</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{21} )</td>
<td>0.5</td>
<td>Bias: - - -</td>
<td>-0.293</td>
<td>-0.001</td>
<td>-0.404</td>
<td>-0.001</td>
<td>-0.342</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE: - - -</td>
<td>0.401</td>
<td>0.081</td>
<td>0.412</td>
<td>0.081</td>
<td>0.358</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{22} )</td>
<td>0</td>
<td>Bias: - - -</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.001</td>
<td>1.165</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE: - - -</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.056</td>
<td>1.758</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \pi_{w} )</td>
<td>0</td>
<td>Bias: - -</td>
<td>0.007</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.165</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE: - -</td>
<td>0.077</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.056</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \pi_{w}^2 )</td>
<td>1</td>
<td>Bias: - -</td>
<td>-0.004</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE: - -</td>
<td>0.085</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: * denotes a QML estimator with the correctly specified variance equation. The reported statistics are calculated from 5,000 replications of the experiment, using samples of 500 observations and the following formulae: \( \text{Bias} = \frac{1}{5,000} \sum_{i=1}^{5,000} (\hat{\theta}_i - \theta) \), \( \text{RMSE} = \sqrt{\frac{1}{5,000} \sum_{i=1}^{5,000} (\hat{\theta}_i - \theta)^2} \).
Table 3. Descriptive Statistics of Data on August 30, 2012

<table>
<thead>
<tr>
<th></th>
<th>Price</th>
<th>Volume</th>
<th>EBITDA</th>
<th>Market</th>
<th>Book to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>33.39</td>
<td>241.19</td>
<td>237.63</td>
<td>1.25</td>
<td>0.57</td>
</tr>
<tr>
<td>Stddev</td>
<td>66.45</td>
<td>580.36</td>
<td>896.99</td>
<td>0.49</td>
<td>0.56</td>
</tr>
<tr>
<td>Min</td>
<td>0.26</td>
<td>0.09</td>
<td>-216.43</td>
<td>-0.06</td>
<td>-3.18</td>
</tr>
<tr>
<td>L5%</td>
<td>3.49</td>
<td>4.72</td>
<td>-6.74</td>
<td>0.49</td>
<td>0.07</td>
</tr>
<tr>
<td>Q1</td>
<td>10.71</td>
<td>24.02</td>
<td>9.63</td>
<td>0.90</td>
<td>0.27</td>
</tr>
<tr>
<td>Median</td>
<td>21.82</td>
<td>77.19</td>
<td>39.62</td>
<td>1.22</td>
<td>0.48</td>
</tr>
<tr>
<td>Q3</td>
<td>41.24</td>
<td>216.83</td>
<td>135.73</td>
<td>1.57</td>
<td>0.77</td>
</tr>
<tr>
<td>U5%</td>
<td>85.08</td>
<td>913.73</td>
<td>941.77</td>
<td>2.10</td>
<td>1.32</td>
</tr>
<tr>
<td>Max</td>
<td>2118.75</td>
<td>7503.58</td>
<td>15522.00</td>
<td>3.14</td>
<td>8.36</td>
</tr>
<tr>
<td>Skewness</td>
<td>19.10</td>
<td>6.84</td>
<td>9.87</td>
<td>0.36</td>
<td>3.57</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>536.35</td>
<td>63.62</td>
<td>126.95</td>
<td>3.10</td>
<td>47.10</td>
</tr>
</tbody>
</table>

Here prices and EBITDA are in the U.S. dollars and volumes are in thousand shares. The L5% and U5% entries are the lower and upper 5% quantiles respectively. The Q1 and Q3 entries are the 25% and 75% quantiles respectively. There are 1902 observations.
Table 4: Parameter Estimates for August 30, 2012

<table>
<thead>
<tr>
<th></th>
<th>Supply Equation</th>
<th>Demand Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>logPrice</td>
<td>logVolume</td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.755</td>
<td>15.088</td>
</tr>
<tr>
<td></td>
<td>(0.964)</td>
<td>(0.729)</td>
</tr>
<tr>
<td>logVolume</td>
<td>0.426</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.086)</td>
<td></td>
</tr>
<tr>
<td>logPrice</td>
<td></td>
<td>-1.149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.220)</td>
</tr>
<tr>
<td>EBITDA</td>
<td>-0.133</td>
<td>4.706</td>
</tr>
<tr>
<td></td>
<td>(0.093)</td>
<td>(0.672)</td>
</tr>
<tr>
<td>BETA</td>
<td>-0.271</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>Book-Market</td>
<td>-0.291</td>
<td>-0.541</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.099)</td>
</tr>
<tr>
<td></td>
<td>0.082</td>
<td>1.007</td>
</tr>
<tr>
<td></td>
<td>(0.148)</td>
<td>(0.133)</td>
</tr>
<tr>
<td>EBITDA</td>
<td>-0.202</td>
<td>1.024</td>
</tr>
<tr>
<td></td>
<td>(0.100)</td>
<td>(0.128)</td>
</tr>
<tr>
<td>BETA</td>
<td>0.069</td>
<td>-0.056</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>Book-Market</td>
<td>-0.162</td>
<td>-0.129</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.049)</td>
</tr>
<tr>
<td>EBITDA(^2)</td>
<td>0.022</td>
<td>-0.033</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>BETA(^2)</td>
<td>-0.002</td>
<td>0.073</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Book-Market(^2)</td>
<td>0.032</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Reduced-form error BP-test (df=3)</td>
<td>46.202</td>
<td>141.731</td>
</tr>
<tr>
<td>Standardised structural error BP-test (df=3)</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The standard errors of estimates are in the parentheses. The standard errors are computed from the robust sandwich form variance matrix of the estimates. The BP-test is the Breusch-Pagan heteroskedasticity test statistic.
Table 5: Parameter Estimates for August 30, 2012
with Linear Conditional Variances

<table>
<thead>
<tr>
<th>Mean Equations</th>
<th>Supply Equation logPrice</th>
<th>Demand Equation logVolume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.547</td>
<td>15.472</td>
</tr>
<tr>
<td></td>
<td>(0.723)</td>
<td>(0.538)</td>
</tr>
<tr>
<td>Volume</td>
<td>0.408</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td></td>
<td>-1.107</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.165)</td>
</tr>
<tr>
<td>EBITDA</td>
<td>-0.106</td>
<td>6.252</td>
</tr>
<tr>
<td></td>
<td>(0.067)</td>
<td>(0.692)</td>
</tr>
<tr>
<td>BETA</td>
<td>-0.274</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>Book-Market</td>
<td>-0.289</td>
<td>-0.443</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.143)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance Equation</th>
<th>Supply Equation logPrice</th>
<th>Demand Equation logVolume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.089</td>
<td>0.934</td>
</tr>
<tr>
<td></td>
<td>(0.079)</td>
<td>(0.273)</td>
</tr>
<tr>
<td>EBITDA</td>
<td>-0.180</td>
<td>-0.082</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>BETA</td>
<td>0.052</td>
<td>-0.303</td>
</tr>
<tr>
<td></td>
<td>(0.050)</td>
<td>(0.135)</td>
</tr>
<tr>
<td>Book-Market</td>
<td>-0.371</td>
<td>-0.955</td>
</tr>
<tr>
<td></td>
<td>(0.090)</td>
<td>(0.087)</td>
</tr>
<tr>
<td>“α_{k2}”</td>
<td>0.307</td>
<td>-4.197</td>
</tr>
<tr>
<td></td>
<td>(0.112)</td>
<td>(0.546)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tests</th>
<th>Supply Equation logPrice</th>
<th>Demand Equation logVolume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced-form error BP-test (df=3)</td>
<td>46.202</td>
<td>141.731</td>
</tr>
<tr>
<td>Standardised structural error BP-test (df=3)</td>
<td>1.707</td>
<td>28.067</td>
</tr>
</tbody>
</table>

The standard errors of estimates are in the parentheses. The standard errors are computed from the robust sandwich form variance matrix of the estimates. The BP-test is the Breusch-Pagan heteroskedasticity test statistic.
Table 6: Parameter Estimates for August 30, 2012
with Logarithms of EBITDA and Book-Market

<table>
<thead>
<tr>
<th></th>
<th>Supply Equation</th>
<th>Demand Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>logPrice</td>
<td>logVolume</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.986</td>
<td>10.663</td>
</tr>
<tr>
<td></td>
<td>(0.301)</td>
<td>(0.271)</td>
</tr>
<tr>
<td>logVolume</td>
<td>0.124</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td></td>
</tr>
<tr>
<td>logPrice</td>
<td></td>
<td>-.820</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.073)</td>
</tr>
<tr>
<td>logEBITDA</td>
<td>0.169</td>
<td>0.872</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>BETA</td>
<td>-0.347</td>
<td>0.178</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.049)</td>
</tr>
<tr>
<td>logBook-Market</td>
<td>-1.168</td>
<td>-1.290</td>
</tr>
<tr>
<td></td>
<td>(0.101)</td>
<td>(0.172)</td>
</tr>
<tr>
<td>Mean Equations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.086</td>
<td>1.561</td>
</tr>
<tr>
<td></td>
<td>(0.217)</td>
<td>(0.196)</td>
</tr>
<tr>
<td>logEBITDA</td>
<td>-0.090</td>
<td>-0.457</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>BETA</td>
<td>0.209</td>
<td>0.229</td>
</tr>
<tr>
<td></td>
<td>(0.314)</td>
<td>(0.282)</td>
</tr>
<tr>
<td>logBook-Market</td>
<td>-1.407</td>
<td>-1.141</td>
</tr>
<tr>
<td></td>
<td>(0.213)</td>
<td>(0.206)</td>
</tr>
<tr>
<td>logEBITDA(^2)</td>
<td>0.008</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>BETA(^2)</td>
<td>-0.049</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>(0.118)</td>
<td>(0.104)</td>
</tr>
<tr>
<td>logBook-Market(^2)</td>
<td>0.920</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td>(0.156)</td>
<td>(0.194)</td>
</tr>
<tr>
<td>Variance Equations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduced-form error BP-test (df=3)</td>
<td>101.839</td>
<td>153.069</td>
</tr>
<tr>
<td>Tests</td>
<td>Standardised structural error BP-test (df=3)</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The standard errors of estimates are in the parentheses. The standard errors are computed from the robust sandwich form variance matrix of the estimates. The BP-test is the Breusch-Pagan heteroskedasticity test statistic. As EBITDA and Book-Market contain negative values, we define logEBITDA = log(1+|EBITDA|) for non-negative EBITDA and −log(1+|EBITDA|) for negative EBITDA. The same definition applies to logBook-Market.
Figure 1: Average demand and supply elasticities for Russell 3000 companies

log(Price) vs. log(Volume)

- **August 29, 2012**
- **August 30, 2012**
- **August 31, 2012**