GMM-based inference in the AR(1) panel data model for parameter values where local identification fails

Edith Madsen

Centre for Applied Microeconometrics (CAM)
Department of Economics, University of Copenhagen, Denmark

April 2009
PRELIMINARY

Abstract

We are interested in making inference on the AR parameter in an AR(1) panel data model when the time-series dimension is fixed and the cross-section dimension is large. We consider a GMM estimator based on the second order moments of the observed variables. It turns out that when the AR parameter equals unity and certain restrictions apply to the other parameters in the model then local identification fails. We show that the parameters and in particular the AR parameter can still be estimated consistently but at a slower rate than usual. Also we derive the asymptotic distribution of the parameter estimators which turns out to be non-standard. The properties of this estimator of the AR parameter are very different from those of the widely used Arellano-Bond GMM estimator which is obtained by taking first-differences of the variables and then use lagged levels as instruments. It is well-known that this estimator performs poorly when the AR parameter is close to unity. The reason for this difference between the two estimators is that with the Arellano-Bond moment conditions global identification fails for the specific values of the parameters considered here whereas with the non-linear GMM estimator that uses all information from the second order moments only local identification fails. We illustrate the results in a simulation study.

Keywords: AR(1) panel data model, GMM estimation, local identification failure, rate of convergence, non-standard limiting distribution

1 Introduction

We are interested in making inference on the parameters in the simple AR(1) panel data model. This type of model is used to distinguish between two types of persistency often found in economic variables at the individual level such as incomes and wages. Within the framework of this model one source of persistency is attributed to the AR parameter which is common across all cross-section units and it describes how persistent unanticipated shocks to the variables of interest are. The other source of persistency is attributed to unobserved individual-specific effects and reflects differences that remains constant over time for specific cross-section units. Within this type of model the focus has mainly been on estimating the AR parameter whereas the individual-specific effects or the parameters describing specific features of their distribution are treated as nuisance parameters. One of the most widely used estimators of the AR parameter is the Arellano-Bond GMM estimator which is obtained by taking first-differences of the variables in the equation in order to eliminate the individual-specific effects and then use lagged levels of the variable as instruments. This estimator has several advantages. First of all, it is based on moment conditions that are linear in the AR parameter and consequently it is straight forward to calculate. In addition, it only requires very basic assumptions about the different terms in the AR(1) panel data model. In particular, it does not require that the exact relationship between the initial values and the individual-specific effects is specified. However, it is also well-known that this estimator sometimes suffers from a weak instrument problem when the AR parameter is close to unity.

The contribution of this paper is to get a better understanding of when there is a problem with identification (weak instruments in the linear case) and to show that the very basic assumptions implies moment conditions that do contain information about the AR parameter also when this is not the case for the linear Arellano-Bond moment conditions. Using all of the moment conditions gives a simple non-linear GMM estimator and we derive the asymptotic properties of this estimator. We will consider micro-panels where the cross-section dimension N is large and the time-series dimension T is relatively small. This means that the asymptotic approximations to the statistics of interest will be based on letting $N \to \infty$ while T is fixed. Under the same basic assumptions Ahn & Schmidt (1995, 1997) obtain the Arellano-Bond moments conditions and additional moment conditions that are quadratic in the AR parameter. These moment conditions do not depend on the remaining parameters in the model and they are obtained by transforming the original set of moment conditions. The approach in this paper is a somewhat different as we will use the full set of moment conditions and obtain estimators of all parameters used in the specification of the model.

The findings in this paper are based on a result from the statistical literature. Rotnitzky, Cox, Bottai & Robins (2000) consider a likelihood-based inference procedure when the standard assumptions about identification are not satisfied. The application to the GMM-based framework is somewhat similar but there are also some differences. We are able to derive the limiting distribution of the GMM estimator for the particular parameter values where local identification fails. It turns out that it is necessary to consider a Taylor expansion of the GMM objective function of higher order than usual in order to be able to explain its behavior. In particular, we find that the GMM estimator of the AR parameter is $N^{1/4}$-consistent and has a limiting distribution which is a non-standard distribution. In a simulation study we show that this limiting distribution provides a good approximation to the actual distribution of the parameter estimators.

Finally, the findings in Stock & Wright (2000) about the properties of GMM estimators under weak instruments are interesting in relation to the results of this paper. Stock & Wright (2000) consider a general framework with moment conditions that are non-linear in the parameters but linear in the instruments under the assumption that the moment conditions provide only weak identification of one of the parameters. In this case they find that the GMM estimator of the parameter that is only weakly identified is inconsistent whereas the GMM estimator of the remaining parameters is consistent at the usual rate but with a non-standard limiting distribution. In their paper the limiting distributions of the GMM estimators are expressed as being the distribution of the values at which a non-central chi-squared process is minimized. Here in this paper, the moment conditions provide identification of all parameters even though the assumption about first order local identification is not satisfied which implies that the convergence rate is slower than usual. In addition we provide the exact limiting distribution.

2 The model

We consider the AR(1) panel data model with individual-specific effects

$$y_{it} = \rho y_{it-1} + \eta_i + \varepsilon_{it} \quad \text{for } i = 1, ..., N \text{ and } t = 1, ..., T \tag{1}$$

For notational convenience it is assumed that y_{i0} is observed such that the number of observations over time equals $T+1$. By repeated substitution in equation (1) we find that

$$y_{it} = \rho^t y_{i0} + \left(\sum_{s=0}^{t-1} \rho^s\right) \eta_i + \sum_{s=1}^{t} \rho^{t-s} \varepsilon_{is} \quad \text{for } t = 1, 2, ... \tag{2}$$

The following assumptions are imposed:

Assumption 1

$(y_{i0}, \eta_i, \varepsilon_{i1}, ..., \varepsilon_{iT})$ is iid across i and has finite fourth order moments
$E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}^2) = \sigma_\varepsilon^2$ and $E(\varepsilon_{is}\varepsilon_{it}) = 0$ for $s \neq t$ and $s, t = 1, ..., T$
$E(\eta_i) = 0$ and $E(\eta_i^2) = \sigma_\eta^2$
$E(y_{i0}) = 0$ and $E(y_{i0}^2) = \sigma_0^2$
$E(\varepsilon_{it}\eta_i) = 0$ and $E(\varepsilon_{it} y_{i0}) = 0$ for $t = 1, ..., T$
$E(y_{i0}\eta_i) = \sigma_{0\eta}^2$

These assumptions are very general and in particular they do not require that the exact relationship between the initial values y_{i0} and the individual-specific effects η_i is specified. The assumptions imply that the observed variables $y_i = (y_{i0}, ..., y_{iT})'$ are iid across i with first order moments equal to zero and second order moments that can be obtained by using the expression in (2). In the next section this is used as the basis for the estimation of the parameters in the model. Note that we have assumed that the error terms ε_{it} have homoskedastic variances over time. This assumption could be relaxed in order to allow for heteroskedastic variances over time.

Before we continue let us examine the assumption about mean stationarity. This assumption is often applied when deriving the asymptotic properties of parameter estimators in the simple AR(1) panel data model. It imposes a restriction on the relationship between the initial values and the individual-specific effects such that the covariance between the observed variable y_{it} and the individual-specific effect η_i is the same for all $t = 0, 1, ..., T$. Mean stationarity can be formulated in the following two ways:

$$\begin{aligned} 1) \qquad y_{i0} &= \alpha_i + \varepsilon_{i0} \text{ and } \eta_i = (1-\rho)\,\alpha_i \\ 2) \qquad y_{i0} &= \frac{\eta_i}{1-\rho} + \varepsilon_{i0} \text{ for } \rho \neq 1 \end{aligned}$$

where ε_{i0} is iid across i with $E(\varepsilon_{i0}) = 0$ and $E\left(\varepsilon_{i0}^2\right) = \tau\sigma_\varepsilon^2$ for $\tau \geq 0$. Solving equation (1) recursively and using the assumption about mean stationarity we have

$$1) \qquad y_{it} = \alpha_i + \rho^t\varepsilon_{i0} + \rho^{t-1}\varepsilon_{i1} + ... + \varepsilon_{it} \quad \text{for all } t = 0, 1, 2, ... \tag{3}$$

$$2) \qquad y_{it} = \frac{\eta_i}{1-\rho} + \rho^t\varepsilon_{i0} + \rho^{t-1}\varepsilon_{i1} + ... + \varepsilon_{it} \quad \text{for all } t = 0, 1, 2, ... \tag{4}$$

This implies that (and therefore the term 'mean stationary'):

$$\begin{aligned} 1) \qquad E\left(y_{it}|\,\alpha_i\right) &= \alpha_i \quad \text{for all } t = 0, 1, 2, ... \\ 2) \qquad E\left(y_{it}|\,\eta_i\right) &= \frac{\eta_i}{1-\rho} \quad \text{for all } t = 0, 1, 2, ... \end{aligned}$$

As long as the AR parameter ρ is considered as being fixed the two ways of formulating mean stationarity are equivalent. However, when the value of ρ approaches unity (usually formulated as ρ being local-to-unity, i.e. $\rho = 1 - c/\sqrt{N}$) then the two data generating processes are very different. In general under the formulation in 1) we have $\sigma_\eta^2 = (1-\rho)^2\sigma_\alpha^2$ and $\sigma_{0\eta}^2 = (1-\rho)\,\sigma_\alpha^2$ such that $\sigma_{0\eta}^2(1-\rho) = \sigma_\eta^2$. This means that $\rho = 1$ implies that $\sigma_\eta^2 = 0$ and $\sigma_{0\eta}^2 = 0$. It turns out that under the formulation in 1) the correlation between y_{i0} and η_i is fixed as $\rho \to 1$ and zero when $\rho = 1$ whereas under the formulation in 2) the correlation between y_{i0} and η_i tends to ± 1 as $\rho \to 1$ and in addition the variables $y_{i0}, ..., y_{iT}$ are dominated by the term $\eta_i/(1-\rho)$ and close to being linear dependent as $\rho \to 1$.

From the expressions in (3)-(4) it is clear that the Arellano-Bond GMM estimator where lagged levels are used as instruments for the variables in first-differences suffers from a weak instrument problem under mean stationarity and when ρ is close to unity. Under the formulation in 1) the first-differences Δy_{it} can be written as $\Delta y_{it} = \left(\rho^t - \rho^{t-1}\right) \varepsilon_{i0} + ... + (\rho - 1)\, \varepsilon_{it-1} + \varepsilon_{it}$ whereas the instruments $y_{i0}, ..., y_{it-1}$ depend on $\alpha_i, \varepsilon_{i0}, ..., \varepsilon_{it-1}$. This means that there is a weak instrument problem when ρ is close to unity. Under the formulation in 2) this is even more serious as the instruments are dominated by the individual-specific term $\eta_i / (1-\rho)$ which does not appear in the first-differences. In the existing literature it is common to consider the formulation in 2). For example Blundell & Bond (1998) and Blundell, Bond & Windmeijer (2000) explain the weak instrument problem in the AR(1) panel data model within this framework from simulation studies and recently Hahn, Hausman & Kursteiner (2007) have derived the asymptotic behavior of the Arellano-Bond GMM estimator under the assumption that ρ is local-to-unity.

In the present paper we derive the asymptotic properties of a specific GMM estimator based on the restrictions in Assumption 1 when the observed variables are generated by mean stationary AR(1) processes as formulated in 1) with an AR parameter equal to unity. More specifically, we consider the behavior of a GMM estimator when the true parameter values are $\rho = 1$, $\sigma_\eta^2 = 0$ and $\sigma_{0\eta}^2 = 0$. It is important to emphasize that only the restrictions in Assumption 1 and not the additional restrictions implied by mean stationarity are used in the estimation of the parameters.

3 Estimation

For the moment, we assume that we observe y_{i0}, y_{i1}, y_{i2} for all i (3 observations over time).
The model contains the following 5 parameters: $\rho, \sigma_0^2, \sigma_\eta^2, \sigma_{0\eta}^2, \sigma_\varepsilon^2$

The 5×1 vector of parameters θ is defined as $\theta = \left(\rho, \sigma_0^2, \sigma_\eta^2, \sigma_{0\eta}^2, \sigma_\varepsilon^2\right)'$. We let θ_0 denote the true value of the parameter θ. In the following we are interested in estimation of the parameter θ when its true value equals θ^* with $\theta^* = \left(1, \sigma_0^{2*}, 0, 0, \sigma_\varepsilon^{2*}\right)'$ where σ_0^{2*} and σ_ε^{2*} are the true (unknown) values of the parameters σ_0^2 and σ_ε^2.

By using the expression in (2) we have that

$$
\begin{aligned}
y_{i1} &= \rho y_{i0} + \eta_i + \varepsilon_{i1} \\
y_{i2} &= \rho^2 y_{i0} + (1+\rho)\, \eta_i + \rho \varepsilon_{i1} + \varepsilon_{i2}
\end{aligned}
$$

Under Assumption 1 the model yields the following 6 moment conditions:

$$\begin{aligned}
E\left(y_{i0}^2\right) &= \sigma_0^2 \\
E\left(y_{i0}y_{i1}\right) &= \rho\sigma_0^2 + \sigma_{0\eta}^2 \\
E\left(y_{i0}y_{i2}\right) &= \rho^2\sigma_0^2 + (1+\rho)\,\sigma_{0\eta}^2 \\
E\left(y_{i1}^2\right) &= \rho^2\sigma_0^2 + \sigma_\eta^2 + 2\rho\sigma_{0\eta}^2 + \sigma_\varepsilon^2 \\
E\left(y_{i1}y_{i2}\right) &= \rho^3\sigma_0^2 + (1+\rho)\,\sigma_\eta^2 + \rho\,(1+2\rho)\,\sigma_{0\eta}^2 + \rho\sigma_\varepsilon^2 \\
E\left(y_{i2}^2\right) &= \rho^4\sigma_0^2 + (1+\rho)^2\,\sigma_\eta^2 + 2\rho^2\,(1+\rho)\,\sigma_{0\eta}^2 + \left(1+\rho^2\right)\sigma_\varepsilon^2
\end{aligned}$$

Using vector notation this can be written as

$$E\left[g\left(y_i, \theta_0\right)\right] = 0 \tag{5}$$

where the 6×1 vector $g\left(y_i, \theta\right)$ corresponding to the expressions above is defined in Appendix A. The moment conditions are based on second order moments of the observed variable $y_i = \left(y_{i0}, y_{i1}, y_{i2}\right)'$. We see that for a fixed value of ρ then the expressions on the right hand side in the equations above are linear in the remaining parameters $\left(\sigma_0^2, \sigma_\eta^2, \sigma_{0\eta}^2, \sigma_\varepsilon^2\right)'$. Note that we do not use the additional moment conditions implied by mean stationarity here. The variance of moment function $g\left(y_i, \theta\right)$ evaluated at the true parameter value $\theta = \theta_0$ is given by

$$\operatorname{Var}\left[g\left(y_i, \theta_0\right)\right] = E\left[g\left(y_i, \theta_0\right) g\left(y_i, \theta_0\right)'\right] = \Sigma \tag{6}$$

where Σ is positive definite a 6×6 matrix that depends on the fourth order moments of the observed variables $\left(y_{i0}, y_{i1}, y_{i2}\right)'$.

Ahn & Schmidt (1995, 1997) show that under Assumption 1 and when there are 3 observations over time then the following two moment conditions hold

$$E\left(y_{i0}\left(y_{i2} - \rho y_{i1}\right)\right) = 0 \tag{7}$$

$$E\left(\left(y_{i1} - \rho y_{i0}\right)^2 - \left(y_{i2} - \rho y_{i1}\right)^2\right) = 0 \tag{8}$$

The first is the linear Arellano-Bond moment condition and the other is quadratic in the AR parameter ρ. Note that they do not depend on the variance parameters $\left(\sigma_0^2, \sigma_\eta^2, \sigma_{0\eta}^2, \sigma_\varepsilon^2\right)'$. These two moment conditions consist of a subset of the full set of moment conditions in (5) and therefore there might be a loss of information by only using these two moment conditions. Crepon, Kramarz & Trognon (1997) show that when the usual first order asymptotic approximations apply then the GMM estimator of the AR parameter ρ based on the moment conditions in (5) and the GMM estimator based on the two moments above are asymptotically equivalent. This might not be the case here in this paper since the first order asymptotic approximations are not valid. Therefore we will consider the full set of moment conditions in (5).

3.1 Identification

The condition for global identification is

$$E\left[g\left(y_i,\theta\right)\right]=0 \text{ if and only if } \theta=\theta_0 \tag{9}$$

The condition for first order local identification is

$$\operatorname{rank}\left\{E\left[\frac{\partial g}{\partial\theta}\left(y_i,\theta_0\right)\right]\right\}=5 \tag{10}$$

It turns out that first order local identification fails at $\theta_0=\theta^*$ since

$$\operatorname{rank}\left\{E\left[\frac{\partial g}{\partial\theta}\left(y_i,\theta^*\right)\right]\right\}=\operatorname{rank}\left\{\frac{\partial g}{\partial\theta}\left(y_i,\theta^*\right)\right\}=4 \tag{11}$$

In Appendix A it is shown that the derivatives with respect to $\rho,\sigma_\eta^2,\sigma_{0\eta}^2,\sigma_\varepsilon^2$ are linear dependent. This does not imply that these parameters are not identified when $\theta=\theta^*$ but it does mean that the behavior of the GMM estimator can not be explained by the usual first-order asymptotic approximations. When (11) holds then there is a reparametrization $\psi=\psi\left(\theta\right)$ such that $\psi_1=\rho$ and $\tilde{\psi}=\left(\psi_2,..,\psi_5\right)'$ and where it holds that the expected value of the first order derivative of $g\left(y_i,\psi\right)$ with respect to ψ_1 equals zero at $\psi^*=\psi\left(\theta^*\right)$ and the 6×4 matrix of the expected value of the first order derivatives of $g\left(y_i,\psi\right)$ with respect to $\tilde{\psi}$ has full rank equal to 4 at $\psi^*=\psi\left(\theta^*\right)$. This means that in a neighborhood around ψ^* the moment conditions can be written as

$$E\left[g\left(y_i,\psi\right)\right]=E\left[\frac{\partial g}{\partial\tilde{\psi}}\left(y_i,\psi^*\right)\right]\left(\tilde{\psi}-\tilde{\psi}^*\right)+E\left[\frac{\partial^2 g}{\partial\psi_1^2}\left(y_i,\psi^*\right)/2!\right]\left(\psi_1-\psi_1^*\right)^2 \tag{12}$$

The right hand side in the expression above equals zero if and only if $\psi=\psi^*$ when the following condition holds:

$$E\left[\frac{\partial^2 g}{\partial\psi_1^2}\left(y_i,\psi^*\right)\right]\neq0 \text{ and it is not a linear combination of } E\left[\frac{\partial g}{\partial\tilde{\psi}}\left(y_i,\psi^*\right)\right] \tag{13}$$

In particular, this means that first order local identification is satisfied for the parameters $\left(\psi_1-\psi_1^*\right)^2$ $=\left(\rho-1\right)^2$ and $(\tilde{\psi}-\tilde{\psi}^*)$. It turns out that (13) is satisfied in the case we consider in this paper. An implication of this is that when the true value of θ equals θ^* it is possible to estimate the parameters $\left(\rho-1\right)^2$ and $(\tilde{\psi}-\tilde{\psi}^*)$ consistently at the usual $\sqrt{N}$-rate. This is shown in the next section where we also show that it is possible to estimate the parameter of interest θ consistently at a rate of $N^{1/4}$ when its true value equals θ^*.

Let us briefly have a look at the moment conditions underlying the Arellano-Bond GMM estimator. They are given by

$$E\left[\left(\Delta y_{i2}-\rho\Delta y_{i1}\right)y_{i0}\right]=\rho E\left(y_{i0}^2\right)-\left(1-\rho\right)E\left(y_{i0}y_{i1}\right)+E\left(y_{i0}y_{i2}\right) \tag{14}$$

They only depend on ρ and are linear in this parameter. With moment conditions that are linear in the parameters it holds that the concepts of first order local and global identification are equivalent.

It turns out that identification fails when $\rho = 1$ and $\sigma^2_{0\eta} = 0$ and in particular mean-stationarity and an AR parameter equal to unity lead to lack of identification. In general identification fails when $E\left(y_{i0}^2\right) = E\left(y_{i0}y_{i1}\right) = E\left(y_{i0}y_{i2}\right)$. Note that the Arellano-Bond moment conditions in (14) can be written as linear combinations of the moment conditions in (5). In particular, they are a subset of the moment conditions in (5) meaning that they contain less information about the AR parameter.

3.2 The GMM estimator

We consider the GMM estimator $\hat{\theta}$ that minimizes a quadratic form in the sample moments $\frac{1}{N}\sum_{i=1}^{N} g\left(y_i, \theta\right)$, that is

$$\hat{\theta} = \arg\min_{\theta} Q\left(\theta\right) \tag{15}$$

where

$$Q\left(\theta\right) = \left[\frac{1}{N}\sum_{i=1}^{N} g\left(y_i, \theta\right)\right]' W_N \left[\frac{1}{N}\sum_{i=1}^{N} g\left(y_i, \theta\right)\right] \tag{16}$$

The weight matrix W_N is positive semi-definite and may depend on the data but converges in probability to a positive definite matrix W as $N \to \infty$. In the following we will use the identity matrix as a weight matrix. This is done for simplicity and a different weight matrix might be used in the future. Also it is not clear if the weight matrix that is found to be optimal under the standard assumption about local identification is optimal when this assumption fails.

Since we are first of all interested in the AR parameter ρ and since the moment conditions in (5) for fixed ρ are linear in the other parameters we will consider the concentrated objective function. For a fixed value of ρ we let $\hat{\theta}_2\left(\rho\right)$ denote the GMM estimator of $\theta_2 = \left(\sigma_0^2, \sigma_\eta^2, \sigma_{0\eta}^2, \sigma_\varepsilon^2\right)'$, that is

$$\hat{\theta}_2\left(\rho\right) = \arg\min_{\theta_2} Q\left(\rho, \theta_2\right) \tag{17}$$

In Appendix A we show that for any value of ρ the minimization problem above always have a unique solution and a closed-form expression for $\hat{\theta}_2\left(\rho\right)$ is easily obtained. In addition $\hat{\theta}_2\left(\rho_0\right)$ is a $\sqrt{N}$-consistent estimator of θ_2 if the true value of ρ equals ρ_0. This results from the matrix in (11) having reduced rank equal to 4 (the number of parameters minus 1) and it is crucial for the results in the paper to hold. The concentrated objective function is defined as

$$Q_C\left(\rho\right) = Q\left(\rho, \hat{\theta}_2\left(\rho\right)\right) = g_C\left(\rho\right)' g_C\left(\rho\right) \tag{18}$$

where $g_C\left(\rho\right) = W_N^{1/2}\frac{1}{N}\sum_{i=1}^{N} g\left(y_i\, \rho, \hat{\theta}_2\left(\rho\right)\right)$ is the sample moments evaluated at $\left(\rho, \hat{\theta}_2\left(\rho\right)'\right)'$. Note here that $\hat{\theta}_2\left(\rho\right)$ depends on the weight matrix W_N but for simplicity this is suppressed in the notation. The GMM estimator of ρ can then be expressed as

$$\hat{\rho} = \arg\min_{\rho} Q_C\left(\rho\right) \tag{19}$$

When we use the usual optimal weight matrix, i.e $W = \Sigma^{-1}$, then asymptotically N times the concentrated objective function as a function of ρ is a non-central χ^2-distribution with 2 degrees of freedom (this results from starting with 6 moment conditions and concentrating out 4 parameters). The non-centrality parameter depends on ρ and is given by $N\theta_2^{*\prime}F(1)'\left(I_6 - F(\rho)\left[F(\rho)'F(\rho)\right]^{-1}F(\rho)'\right)$ $F(1)\theta_2^*$ where $F(\rho) = \Sigma^{-1/2}S(\rho)$. In particular, since the non-centrality parameter equals zero at $\rho = 1$ we have that asymptotically $NQ_C(1)$ is a χ^2-distribution with 2 degrees of freedom. It is also clear that the non-centrality parameter is increasing in the number of cross-section observations N. Altogether we might consider $NQ_C(\rho)$ as a non-central chi-squared process indexed by ρ. It is then possible to simulate this process for different values of ρ and obtain the value of ρ where the minimum is obtained. The distribution of these values of ρ would then reflect the asymptotic distribution of the GMM estimator $\hat{\rho}$. This approach is used in Stock & Wright (2000) within a weak instrument framework such that the non-centrality parameter is considered as being constant as N tends to infinity. However, here we will use an approximation of the concentrated objective function in a neighborhood around unity in order to explain the behavior of the GMM estimator $\hat{\rho}$.

3.3 Asymptotic approximations

The derivative of order k of the concentrated objective function is defined as

$$Q_C^{(k)}(\rho) = \frac{d^k Q_C}{d\rho^k}(\rho) \tag{20}$$

Expressions can be found in Appendix A. A Taylor expansion of order k of the concentrated objective function $Q_C(\rho)$ around the true value ρ_0 gives

$$Q_C(\rho) = Q_C(\rho_0) + \sum_{j=1}^{k} Q_C^{(j)}(\rho_0)(\rho-\rho_0)^j / j! + L_k(\rho) \tag{21}$$

where $L_k(\rho) = Q_C^{(k+1)}(\bar{\rho})(\rho-\rho_0)^{k+1} / (k+1)!$ for some $\bar{\rho}$ with $|\bar{\rho}-\rho_0| < |\rho-\rho_0|$. When the "standard regularity conditions" (including first order local identification) are satisfied then the following results concerning derivatives of the concentrated objective function hold

$$\sqrt{N}Q_C^{(1)}(\rho_0) = 2g_C^{(1)}(\rho_0)'\sqrt{N}g_C(\rho_0) \overset{w}{\to} N\left(0, 4G_1'\tilde{\Sigma}G_1\right) \quad \text{as } N \to \infty \tag{22}$$

$$Q_C^{(2)}(\rho_0) \overset{as}{=} 2g_C^{(1)}(\rho_0)'g_C^{(1)}(\rho_0) \overset{P}{\to} 2G_1'G_1 > 0 \quad \text{as } N \to \infty \tag{23}$$

$$Q_C^{(3)}(\rho_0) = O_P(1) \tag{24}$$

where we have used that $\sqrt{N}g_C(\rho_0) \overset{w}{\to} N\left(0,\tilde{\Sigma}\right)$, $g_C^{(1)}(\rho_0) \overset{P}{\to} G_1 \neq 0$ and $g_C^{(k)}(\rho_0) \overset{P}{\to} G_k$ as $N \to \infty$. For $\sqrt{N}(\rho-\rho_0)$ bounded then the expansion in (21) together with the results above imply that the centered (equal to zero at $\rho = \rho_0$) and appropriate scaled concentrated objective function can be approximated by a second order polynomial in $\sqrt{N}(\rho-\rho_0)$ in a neighborhood around ρ_0. We have the following

$$N\left(Q_C(\rho)-Q_C(\rho_0)\right)\overset{as}{=}\sqrt{N}Q_C^{(1)}(\rho_0)\left\{\sqrt{N}(\rho-\rho_0)\right\}+Q_C^{(2)}(\rho_0)/2!\left\{\sqrt{N}(\rho-\rho_0)\right\}^2 \quad (25)$$

When considered as a function of $\sqrt{N}(\rho-\rho_0)$ the right hand side in the expression above has a unique minimum which is asymptotically equivalent to $\sqrt{N}(\hat{\rho}-\rho_0)$ since $\hat{\rho}$ is defined as the value of ρ which minimizes the left hand side in the expression above. This gives

$$\sqrt{N}(\hat{\rho}-\rho_0)\overset{as}{=}\sqrt{N}Q_C^{(1)}(\rho_0)/Q_C^{(2)}(\rho_0)\overset{w}{\to}N\left(0,(G_1'G_1)^{-1}G_1'\tilde{\Sigma}G_1(G_1'G_1)^{-1}\right)\quad \text{as } N\to\infty \quad (26)$$

It means that under the "standard regularity conditions" (including first order local identification) the GMM estimator $\hat{\rho}$ is a $\sqrt{N}$-consistent estimator of ρ and it is asymptotically normally distributed.

In the case considered here in this paper when local identification fails for $\theta_0=\theta^*$ we need a Taylor expansion of higher order in order to derive asymptotic properties of the GMM estimator. We have the following as $N\to\infty$ (for details see Appendix A)

$$N^{3/4}Q_C^{(1)}(1)=2N^{1/4}g_C^{(1)}(1)'\sqrt{N}g_C(1)\overset{P}{\to}0\quad \text{as } N\to\infty \quad (27)$$

$$\sqrt{N}Q_C^{(2)}(1)\overset{as}{=}2g_C^{(2)}(1)'\sqrt{N}g_C(1)\overset{w}{\to}N\left(0,4G_2'\tilde{\Sigma}G_2\right)\quad \text{as } N\to\infty \quad (28)$$

$$N^{1/4}Q_C^{(3)}(1)\overset{P}{\to}0\quad \text{as } N\to\infty \quad (29)$$

$$Q_C^{(4)}(1)\overset{as}{=}6g_C^{(2)}(1)'g_C^{(2)}(1)\overset{P}{\to}6G_2'G_2>0\quad \text{as } N\to\infty \quad (30)$$

$$Q_C^{(5)}(1)\overset{as}{=}20g_C^{(2)}(1)'g_C^{(3)}(1)\overset{P}{\to}20G_2'G_3\quad \text{as } N\to\infty \quad (31)$$

where we have used that $\sqrt{N}g_C(1)\overset{w}{\to}N\left(0,\tilde{\Sigma}\right)$, $N^{1/4}g_C^{(1)}(1)\overset{P}{\to}0$, $g_C^{(2)}(1)\overset{P}{\to}G_2\neq 0$ and $g_C^{(k)}(1)\overset{P}{\to}G_k$ as $N\to\infty$. For $N^{1/4}(\rho-1)$ bounded then the expansion in (21) together with the results above imply that the centered and scaled concentrated objective function can be approximated by a fourth order polynomial in $N^{1/4}(\rho-1)$ in a neighborhood around unity. We have the following

$$\begin{aligned}&N\left(Q_C(\rho)-Q_C(1)\right)\\ =\ &\left\{N^{1/4}(\rho-1)\right\}^2\left(\sqrt{N}Q_C^{(2)}(1)/2!+Q_C^{(4)}(1)/4!\left\{N^{1/4}(\rho-1)\right\}^2\right)+R_N(\rho)\end{aligned} \quad (32)$$

where

$$R_N(\rho)=\omega(\rho)\left(N^{3/4}Q_C^{(1)}(1)+N^{1/4}Q_C^{(3)}(1)/3!\omega(\rho)^2+N^{-1/4}Q_C^{(5)}(\bar{\rho})/5!\omega(\rho)^4\right) \quad (33)$$

$$\omega(\rho)=\left\{N^{1/4}(\rho-1)\right\} \quad (34)$$

It holds that $R_N(\rho)\overset{P}{\to}0$ as $N\to\infty$. Notice that the limiting distributions of the coefficients in the polynomial on the right hand side in equation (32) above are of the same form as in the second order polynomial obtained when the assumption about local identification is satisfied. We have the following

$$N\left(Q_C(\rho)-Q_C(1)\right)\overset{as}{=}\left\{N^{1/4}(\rho-1)\right\}^2\left(\tilde{Z}+1/4G_2'G_2\left\{N^{1/4}(\rho-1)\right\}^2\right) \tag{35}$$

where the right and left hand side in the expression above are asymptotically equivalent and $\tilde{Z}\sim N\left(0,G_2'\tilde{\Sigma}G_2\right)$. In the case where we have a positive value of $\tilde{Z}$ (this will happen with probability 1/2) then the right hand side in the expression above has a global minimum at $\rho=1$. In the case where we have a negative value of $\tilde{Z}$ (this will happen with probability 1/2) then the expression on the right hand side will have a local maximum at $\rho=1$ and two global minima attained at $N^{1/4}(\rho-1)=\pm\sqrt{-\tilde{Z}/\left(1/2G_2'G_2\right)}$. This means that $\hat{\rho}=1$ with probability 1/2 and $\hat{\rho}=1\pm N^{-1/4}\sqrt{-\tilde{Z}/\left(1/2G_2'G_2\right)}$ with probability 1/2. In order to determine which of these values gives the minimum of the objective function we will need to consider the deviation of the scaled objective function and the expression on the right hand side in the expression above evaluated at the two values. Using equation (33) we write $R_N(\rho)=\omega(\rho)T_N(\rho)$ where

$$T_N(\rho)=N^{3/4}Q_C^{(1)}(1)+N^{1/4}Q_C^{(3)}(1)/3!\omega(\rho)^2+N^{-1/4}Q_C^{(5)}(\bar{\rho})/5!\omega(\rho)^4 \tag{36}$$

We see that $T_N(\hat{\rho}_1)=T_N(\hat{\rho}_2)$ for $\hat{\rho}_1=1-\sqrt{-\tilde{Z}/\left(1/2G_2'G_2\right)}/N^{1/4}$ and $\hat{\rho}_2=1+\sqrt{-\tilde{Z}/\left(1/2G_2'G_2\right)}/N^{1/4}$. This means that the sign of $T_N(\hat{\rho})$ will determine at which of the two values the minimum is attained. If the sign of $T_N(\rho)$ is negative (positive) the minimum is attained at $\hat{\rho}_2$ ($\hat{\rho}_1$). This has to be conditional on getting a negative draw of $\sqrt{N}Q_C^{(2)}(1)/2!\overset{as}{=}\tilde{Z}$. So in order to determine where the minimum is attained we have to consider the conditional distribution of $N^{1/4}T_N(\hat{\rho})$ given $\sqrt{N}Q_C^{(2)}(1)/2!$ and determine the probability that $T_N(\hat{\rho})<0$ conditional on $\sqrt{N}Q_C^{(2)}(1)<0$. The difference between the framework considered here in this paper and the one in Rotnitzky, Cox, Bottai & Robins (2000) is that the expected value of the first order derivative of the concentrated moment function at ρ equal to unity equals zero, i.e. $E\left[g_C^{(1)}(1)\right]=0$, while Rotnitzky, Cox, Bottai & Robins (2000) assume that the first order derivative of their objective function (the log-likelihood function) at this point is equal to zero with probability 1. This means that in our framework $\sqrt{N}g_C^{(1)}(1)=O_P(1)$ such that expressions in $N^{1/4}T_N(\hat{\rho})$ involving this term will appear and that complicate things. In fact it turns out that if the usual optimal weight matrix $W_N=\Sigma^{-1}$ is used and if it was the case that $g_C^{(1)}(1)=0$ then we would have that $N^{1/4}T_N(\hat{\rho})$ is asymptotically normal with mean zero and independent of $\sqrt{N}Q_C^{(2)}(1)/2!$ such that the minimum is attained at $\hat{\rho}_1$ and $\hat{\rho}_2$ with equal probability.

The limiting distribution of the GMM estimator of the AR parameter $\hat{\rho}$ is given in the proposition below.

Proposition 1 *Under Assumption 1 and when the true value of θ equals θ^* then the following holds:*

$$N^{1/4}(\hat{\rho}-1)\overset{w}{\to}\mathbf{1}_{(Z_1>0)}\cdot 0+\mathbf{1}_{(Z_1<0)}\cdot(-1)^B\sqrt{-Z_1}\quad as\ N\to\infty \tag{37}$$

where $B \sim Ber\,(0.5)$ *and* $Z_1 \sim N\,(0, \Omega_{11})$ *with* $\Omega_{11} = 4\,(G_2'G_2)^{-1}\,G_2'\tilde{\Sigma}G_2\,(G_2'G_2)^{-1}$. *In addition* B *and* Z_1 *are independent of each other (as explained above this is not shown yet).*

The result shows that $\hat{\rho}$ is a $N^{1/4}$-consistent estimator of the AR parameter and that its asymptotic distribution is a mixture with 50% of the probability mass at the true value of unity and 25% each at unity plus/minus the square root of a half-normal distribution. This gives a distribution which is symmetric around the true value of unity with 3 modes of which one is at unity and where the two other modes will become closer to unity as the sample size N increases. Also note that the result implies that the parameter $(\rho - 1)^2$ can be estimated consistently at the usual $\sqrt{N}$-rate and that the limiting distribution of $\sqrt{N}\,(\hat{\rho} - 1)^2$ is a mixture with 50% of the probability mass at the true value which equals 0 and 50% distributed according to the positive half-normal of Z_1. This result is similar to what is found for the behavior of parameter estimators when the true value of the parameter is on the boundary of the parameter space, see Geyer (1994) and Andrews (1999).

The limiting distribution of the GMM estimator of the remaining parameters θ_2 follows by the results in the proposition below.

Proposition 2 *Under Assumption 1 and when the true value of* θ *equals* θ^* *then the following holds as* $N \to \infty$*:*

$$\sqrt{N}\left(\hat{\theta}_2\,(\hat{\rho}) - \theta_2^* - H_1\,(\hat{\rho} - 1) - H_2\,(\hat{\rho} - 1)^2\right) \xrightarrow{w} Z_2 \tag{38}$$

$$\sqrt{N}\left(\hat{\theta}_2\,(\hat{\rho}) - \theta_2^* - H_1\,(\hat{\rho} - 1) - H_2\,(\hat{\rho} - 1)^2\right)\Big|\, N^{1/4}\,(\hat{\rho} - 1) \xrightarrow{w} (\,Z_2|\,Z_1) \tag{39}$$

where

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}\right) \tag{40}$$

with

$$\begin{aligned} \Omega_{11} &= 4\,(G_2'G_2)^{-1}\,G_2'\tilde{\Sigma}G_2\,(G_2'G_2)^{-1} \\ \Omega_{22} &= M\,(1)\,W^{1/2}\Sigma W^{1/2}M\,(1) \\ \Omega_{12} &= 2\,(G_2'G_2)^{-1}\,G_2'\tilde{\Sigma}W^{1/2}M\,(1)' \\ M\,(1) &= \left[F\,(1)'\,F\,(1)\right]^{-1}F\,(1)' \end{aligned}$$

$$H_1 = \begin{bmatrix} 0 \\ \sigma_\varepsilon^{2*} \\ \sigma_0^{2*} \\ -\sigma_\varepsilon^{2*} \end{bmatrix} \qquad H_2 = M\,(1)\left(2F^{(1)}\,(1)\,H_1 - F^{(2)}\,(1)\,\theta_2^*\right)$$

The result in (38) shows which possibly non-linear transformations of the parameter vector θ that can be estimated consistently at the usual $\sqrt{N}$-rate when the true value of θ equals

θ^*. The result also gives the marginal asymptotic distributions of the GMM estimator of the parameter θ_2 since it implies that $N^{1/4}\left(\hat{\theta}_2(\hat{\rho})-\theta_2^*\right)\overset{as}{=}H_1(\hat{\rho}-1)$. In particular, this means that $N^{1/4}\left(\hat{\sigma}_0^2-\sigma_0^{2*}\right)\overset{as}{=}0$ whereas the asymptotic distributions of the remaining parameters are similar to the one obtained for the GMM estimator of the AR parameter $\hat{\rho}$, see Proposition 1. In particular, they can be estimated consistently at the rate $N^{1/4}$. The result also shows that when using the usual optimal weight matrix $W_N=\Sigma^{-1}$ then the estimator of the transformation of the parameters that can be estimated $\sqrt{N}$-consistently is asymptotically independent of the estimator $\hat{\rho}$. The result in (39) shows the conditional distribution of $\sqrt{N}\left(\hat{\theta}_2(\hat{\rho})-\theta_2^*\right)$ given the value of $N^{1/4}(\hat{\rho}-1)$. In general we find that conditional on Z_1 (see Proposition 1) then $\hat{\theta}_2(\hat{\rho})$ is asymptotically normal with mean $\theta_2^*+H_1(\hat{\rho}-1)+H_2(\hat{\rho}-1)^2-\Omega_{21}\Omega_{11}^{-1}Z_1$ and variance $\left(\Omega_{22}-\Omega_{21}\Omega_{11}^{-1}\Omega_{12}\right)/N$. Since $N^{1/4}(\hat{\rho}-1)=\pm\sqrt{-Z_1}$ for $Z_1<0$ and $N^{1/4}(\hat{\rho}-1)=0$ for $Z_1>0$ this means that conditional on $(\hat{\rho}-1)$ then with 50% probability $\hat{\theta}_2(\hat{\rho})$ is asymptotically normal with mean $\theta_2^*+E\left(Z_2|\,Z_1<0\right)$ and with 50% probability $\hat{\theta}_2(\hat{\rho})$ is asymptotically normal with mean $\theta_2^*+H_1(\hat{\rho}-1)+\left(H_2+\Omega_{21}\Omega_{11}^{-1}\right)(\hat{\rho}-1)^2$. When the usual optimal weight matrix is used, that is $W_N=\Sigma^{-1}$, then conditional on $(\hat{\rho}-1)^2$ then with 50% probability $\hat{\theta}_2(\hat{\rho})$ is asymptotically normal with mean θ_2^* and with 50% probability $\hat{\theta}_2(\hat{\rho})$ is asymptotically normal with mean $\theta_2^*+H_1(\hat{\rho}-1)+H_2(\hat{\rho}-1)^2$.

4 A simulation study

The results derived in the previous section are now illustrated in a simulation study. We consider the AR(1) panel data model with parameter values where local identification fails, i.e. $\rho=1,\sigma_0^{2*}=2,\sigma_\eta^2=0,\sigma_{0\eta}^2=0,\sigma_\varepsilon^{2*}=1$. We consider 5000 replications of the model:

$$\begin{aligned} y_{it} &= y_{it-1}+\varepsilon_{it} \quad \text{for } i=1,...,N \text{ and } t=1,2 \\ y_{i0} &\sim \text{iid } N\left(0,\sigma_0^{2*}\right) \quad \text{across } i \\ \varepsilon_{it} &\sim \text{iid } N\left(0,\sigma_\varepsilon^{2*}\right) \quad \text{across } i,t \end{aligned}$$

We have that $N=1000$ and $N=2500$. In practice the minimum of the concentrated objective function $Q_C(\rho)$ is found by a grid search in the interval $[0.2;1.8]$. The reason for this is that the function $Q_C(\rho)$ in approximately 50% of the replications have two local minima of which one is the global minimum. We have used the weight matrix $W_N=\Sigma^{-1}$ where $\Sigma=\text{Var}\left[g\left(y_i,\theta^*\right)\right]$.

Overall the outcome of this simulation experiment shows that the asymptotic results derived in the previous section explains the actual behavior of the GMM estimator of the AR parameter well (Figure 1-4). In addition the outcome shows that the joint distribution of the estimators of θ is also very different from what we usually find when they are jointly normally distributed. (Figure 5-8). This follows by the result in Proposition 2 which also explains the strong quadratic relationship (the green plots) between the estimators of for instance ρ and σ_η^2.

Figure 1: Empirical distribution of $\hat{\rho}$ for $N = 1000$

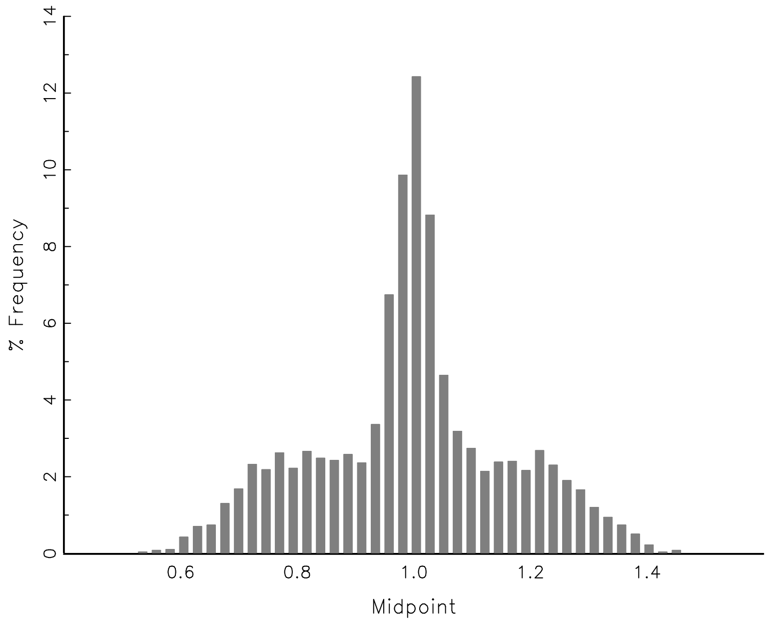


Figure 2: Empirical distribution of $\hat{\rho}$ when $\sqrt{N} Q_C^{(2)}(1) < 0$ for $N = 1000$

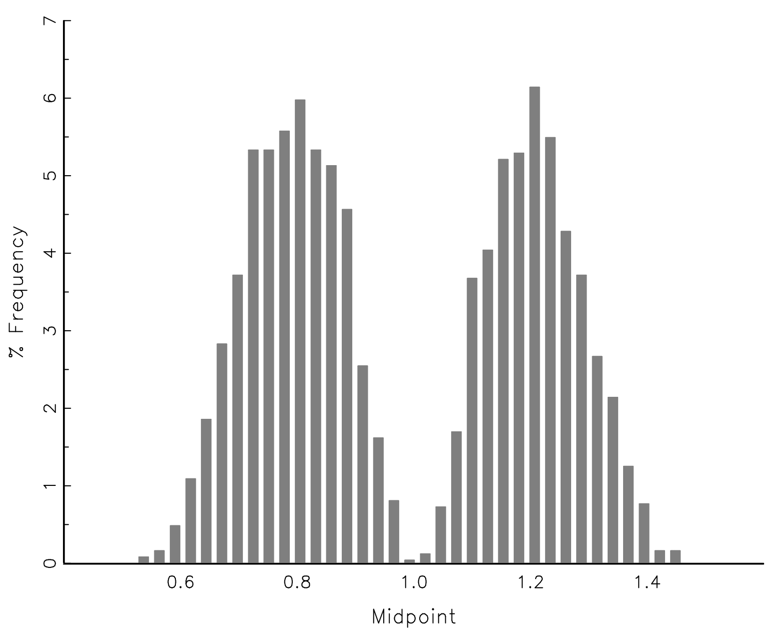


Figure 3: Empirical distribution of $\hat{\rho}$ for $N = 2500$

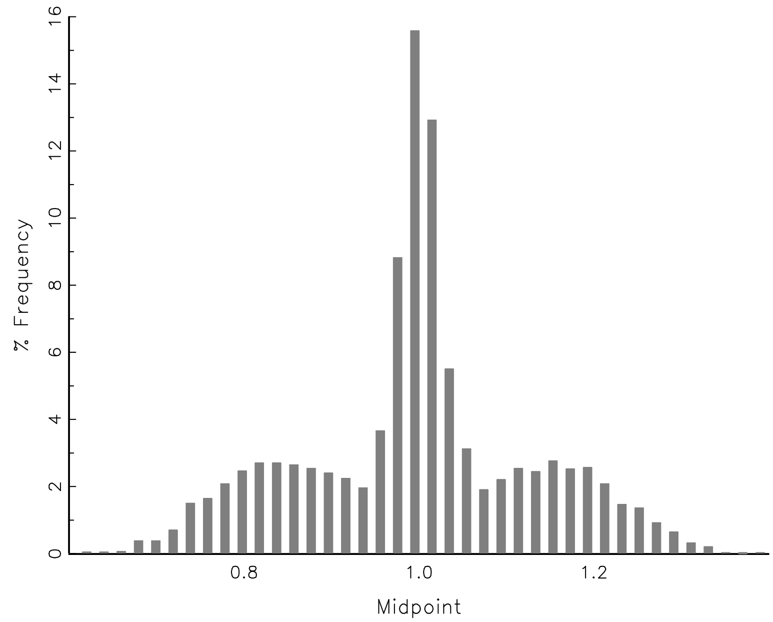


Figure 4: Empirical distribution of $\hat{\rho}$ when $\sqrt{N}Q_C^{(2)}(1) < 0$ for $N = 2500$

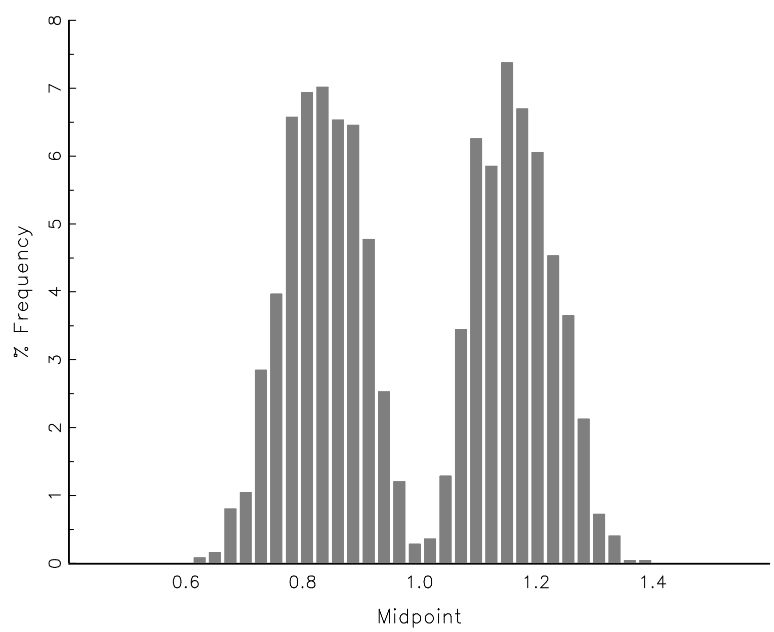


Figure 5: Plot of $(\hat{\rho}, \hat{\sigma}_0^2)$ for $N = 2500$. The red lines are the true parameter values

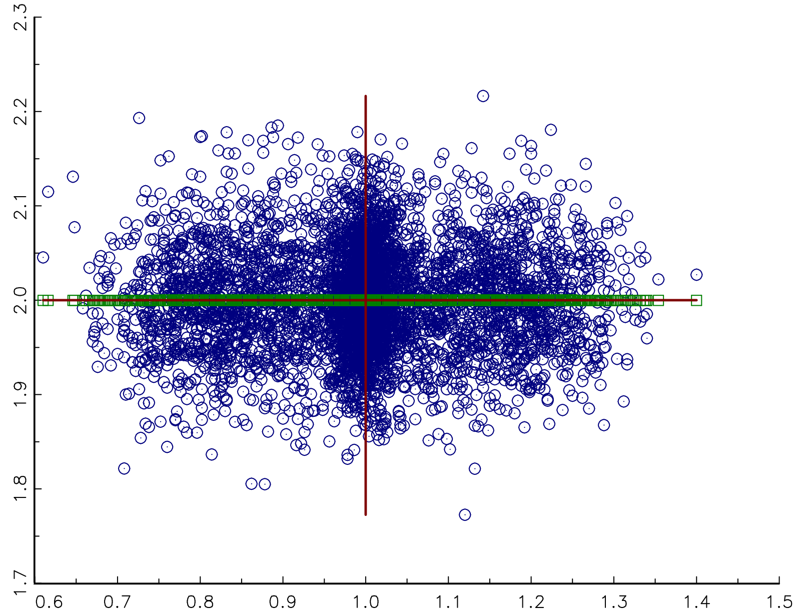


Figure 6: Plot of $(\hat{\rho}, \hat{\sigma}_\eta^2)$ for $N = 2500$. The red lines are the true parameter values

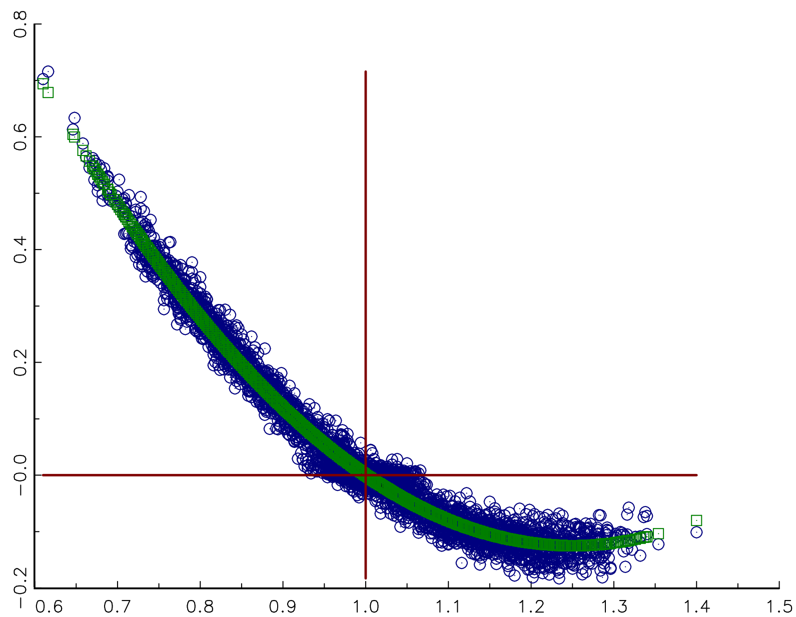


Figure 7: Plot of $(\hat{\rho}, \hat{\sigma}^2_{0\eta})$ for $N = 2500$. The red lines are the true parameter values

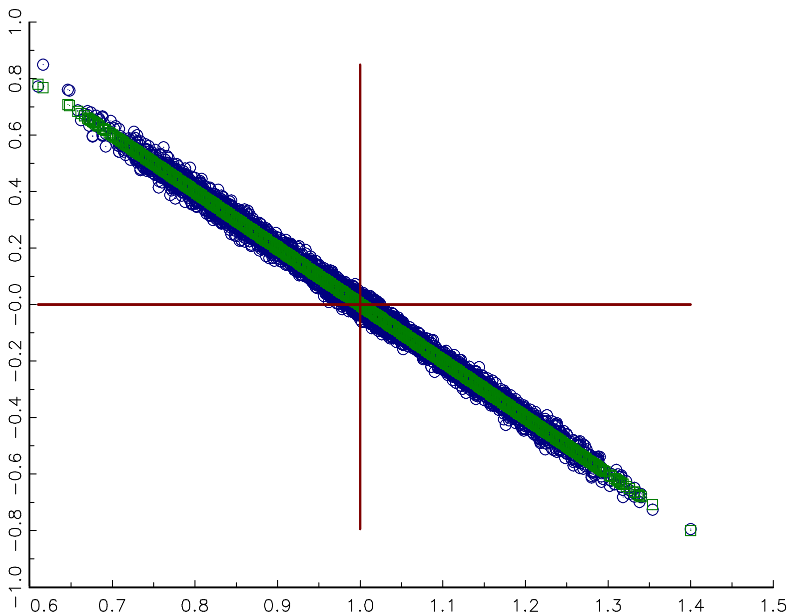
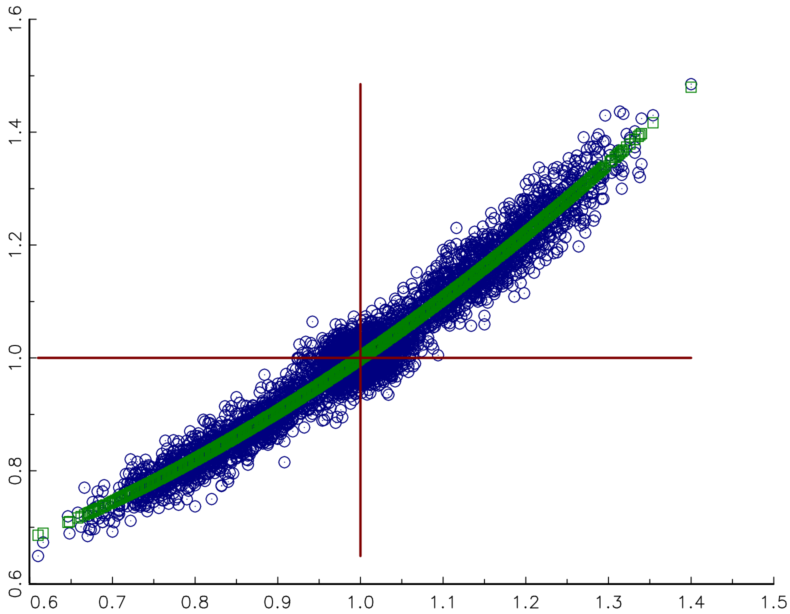


Figure 8: Plot of $(\hat{\rho}, \hat{\sigma}^2_{\varepsilon})$ for $N = 2500$. The red lines are the true parameter values



Appendix A

In this appendix the notation "$X_N \overset{as}{=} Y_N$" means that X_N and Y_N are asymptotically equivalent as $N \to \infty$ such that $X_N - Y_N \overset{P}{\to} 0$ as $N \to \infty$.

The moment conditions:

The 6×1 vector $g(y_i, \theta)$ consists of the following elements

$$\begin{aligned}
g_1(y_i,\theta) &= y_{i0}^2 - \sigma_0^2 \\
g_2(y_i,\theta) &= y_{i0}y_{i1} - \left(\rho\sigma_0^2 + \sigma_{0\eta}^2\right) \\
g_3(y_i,\theta) &= y_{i0}y_{i2} - \left(\rho^2\sigma_0^2 + (1+\rho)\,\sigma_{0\eta}^2\right) \\
g_4(y_i,\theta) &= y_{i1}^2 - \left(\rho^2\sigma_0^2 + \sigma_\eta^2 + 2\rho\sigma_{0\eta}^2 + \sigma_\varepsilon^2\right) \\
g_5(y_i,\theta) &= y_{i1}y_{i2} - \left(\rho^3\sigma_0^2 + (1+\rho)\,\sigma_\eta^2 + \rho\,(1+2\rho)\,\sigma_{0\eta}^2 + \rho\sigma_\varepsilon^2\right) \\
g_6(y_i,\theta) &= y_{i2}^2 - \left(\rho^4\sigma_0^2 + (1+\rho)^2\,\sigma_\eta^2 + 2\rho^2\,(1+\rho)\,\sigma_{0\eta}^2 + \left(1+\rho^2\right)\sigma_\varepsilon^2\right)
\end{aligned}$$

Note that they can be written as

$$g(y_i,\theta) = \mathrm{vech}\left(y_iy_i'\right) - S(\rho)\,\theta_2 \tag{41}$$

where $\theta_2 = \left(\sigma_0^2, \sigma_\eta^2, \sigma_{0\eta}^2, \sigma_\varepsilon^2\right)'$ and the 6×4 matrix $S(\rho)$ is defined as

$$S(\rho) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \rho & 0 & 1 & 0 \\ \rho^2 & 0 & (1+\rho) & 0 \\ \rho^2 & 1 & 2\rho & 1 \\ \rho^3 & (1+\rho) & \rho\,(1+2\rho) & \rho \\ \rho^4 & (1+\rho)^2 & 2\rho^2\,(1+\rho) & \left(1+\rho^2\right) \end{bmatrix} \tag{42}$$

We have that

$$\frac{\partial g}{\partial \theta}(y_i,\theta) = -\left[\begin{array}{ccc} S^{(1)}(\rho)\,\theta_2 & \vdots & S(\rho) \end{array}\right] \tag{43}$$

where $S^{(1)}(\rho)$ denotes the first order derivative of $S(\rho)$ and is given by

$$S^{(1)}(\rho) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 2\rho & 0 & 1 & 0 \\ 2\rho & 0 & 2 & 0 \\ 3\rho^2 & 1 & 1+4\rho & 1 \\ 4\rho^3 & 2\,(1+\rho) & 4\rho+6\rho^2 & 2\rho \end{bmatrix} \tag{44}$$

We see that $S^{(1)}(1)\,\theta_2^* = S(1)\,q$ with $q = \left(0, \sigma_\varepsilon^{2*}, \sigma_0^{2*}, -\sigma_\varepsilon^{2*}\right)'$ such that the matrix in (43) has reduced rank equal to 4 when $\theta = \theta^* = \left(1, \sigma_0^{2*}, 0, 0, \sigma_\varepsilon^{2*}\right)$.

We have

$$E\left[g\left(y_i,\theta_0\right)\right] = 0 \tag{45}$$

$$E\left[g\left(y_i,\theta_0\right)g\left(y_i,\theta_0\right)'\right] = \Sigma \quad \text{(positive definite)} \tag{46}$$

such that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} g\left(y_i,\theta_0\right) \xrightarrow{w} N\left(0,\Sigma\right) \quad \text{as } N\to\infty \tag{47}$$

$$\frac{1}{N}\sum_{i=1}^{N}\left(g\left(y_i,\theta_0\right)-\frac{1}{N}\sum_{i=1}^{N}g\left(y_i,\theta_0\right)\right)\left(g\left(y_i,\theta_0\right)-\frac{1}{N}\sum_{i=1}^{N}g\left(y_i,\theta_0\right)\right)' \xrightarrow{P} \Sigma \quad \text{as } N\to\infty \tag{48}$$

The estimator $\hat{\theta}_2\left(\rho\right)$:

Using the expression in (43) we have that

$$\frac{\partial Q}{\partial\theta_2}\left(\theta\right) = -2\left[\frac{1}{N}\sum_{i=1}^{N}g\left(y_i,\theta\right)\right]' W_N S\left(\rho\right) \tag{49}$$

The FOC for the optimization problem in (17) give

$$\left[\frac{1}{N}\sum_{i=1}^{N}\operatorname{vech}\left(y_i y_i'\right) - S\left(\rho\right)\hat{\theta}_2\left(\rho\right)\right]' W_N S\left(\rho\right) = 0 \tag{50}$$

The FOC are necessary and sufficient since $S\left(\rho\right)' W_N S\left(\rho\right) \stackrel{as}{=} S\left(\rho\right)' W S\left(\rho\right) > 0$ for all values of ρ since W is positive definite and $S\left(\rho\right)$ has full rank equal to 4 for all values of ρ. This implies that

$$\hat{\theta}_2\left(\rho\right) = \left[S\left(\rho\right)' W_N S\left(\rho\right)\right]^{-1} S\left(\rho\right)' W_N \frac{1}{N}\sum_{i=1}^{N}\operatorname{vech}\left(y_i y_i'\right) \tag{51}$$

The concentrated objective function $Q_C\left(\rho\right)$ and its derivatives:

We define the concentrated sample moment function

$$g_C\left(\rho\right) = W_N^{1/2}\frac{1}{N}\sum_{i=1}^{N} g\left(y_i,\rho,\hat{\theta}_2\left(\rho\right)\right) = W_N^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\operatorname{vech}\left(y_i y_i'\right) - S\left(\rho\right)\hat{\theta}_2\left(\rho\right)\right) \tag{52}$$

The concentrated objective function can be written as

$$Q_C\left(\rho\right) = g_C\left(\rho\right)' g_C\left(\rho\right) \tag{53}$$

Using this we obtain the following expressions for derivatives of Q_C

$$\begin{aligned} Q_C^{(1)}(\rho) &= 2g_C(\rho)' g_C^{(1)}(\rho) && (54)\\ Q_C^{(2)}(\rho) &= 2g_C(\rho)' g_C^{(2)}(\rho) + 2g_C^{(1)}(\rho)' g_C^{(1)}(\rho) && (55)\\ Q_C^{(3)}(\rho) &= 2g_C(\rho)' g_C^{(3)}(\rho) + 6g_C^{(1)}(\rho)' g_C^{(2)}(\rho) && (56)\\ Q_C^{(4)}(\rho) &= 2g_C(\rho)' g_C^{(4)}(\rho) + 8g_C^{(1)}(\rho)' g_C^{(3)}(\rho) + 6g_C^{(2)}(\rho)' g_C^{(2)}(\rho) && (57)\\ Q_C^{(5)}(\rho) &= 2g_C(\rho)' g_C^{(5)}(\rho) + 10g_C^{(1)}(\rho)' g_C^{(4)}(\rho) + 20g_C^{(2)}(\rho)' g_C^{(3)}(\rho) && (58) \end{aligned}$$

Inserting the expression for $\hat{\theta}_2(\rho)$ in (51) in equation (52) we have

$$g_C(\rho) = \left(I_6 - F_N(\rho)\left[F_N(\rho)' F_N(\rho)\right]^{-1} F_N(\rho)'\right) W_N^{1/2} \frac{1}{N}\sum_{i=1}^{N} \text{vech}(y_i' y_i) \tag{59}$$

where $F_N(\rho) = W_N^{1/2} S(\rho)$. Note that $\left(I_6 - F_N(\rho)\left[F_N(\rho)' F_N(\rho)\right]^{-1} F_N(\rho)'\right) W_N^{1/2} S(\rho)\theta_2 = 0$ such that when the true value of θ equals θ_0 then the concentrated moment function evaluated at the true value ρ_0 can be written as

$$g_C(\rho_0) = \left(I_6 - F_N(\rho_0)\left[F_N(\rho_0)' F_N(\rho_0)\right]^{-1} F_N(\rho_0)'\right) W_N^{1/2} \frac{1}{N}\sum_{i=1}^{N} g(y_i, \theta_0) \tag{60}$$

The Lemma below gives the limiting distributions of $\sqrt{N} g_C(1)$ and $\sqrt{N}\hat{\theta}_2(1)$ when the true value of θ equal θ^*.

Lemma 3 *When $\theta = \theta^*$ the following holds as $N \to \infty$*

$$M_N(1) W_N^{1/2} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} g(y_i, \theta^*) \xrightarrow{w} N\left(0, M(1) W^{1/2} \Sigma W^{1/2} M(1)'\right) \tag{61}$$

$$(I_6 - P_N(1)) W_N^{1/2} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} g(y_i, \theta^*) \xrightarrow{w} N\left(0, (I_6 - P(1)) W^{1/2} \Sigma W^{1/2} (I_6 - P(1))\right) \tag{62}$$

where

$$\begin{aligned} F_N(\rho) &= W_N^{1/2} S(\rho) = W_N^{1/2} \frac{\partial g(y_i, \theta)}{\partial \theta_2}\\ F(\rho) &= W^{1/2} S(\rho)\\ M_N(\rho) &= \left[F_N(\rho)' F_N(\rho)\right]^{-1} F_N(\rho)'\\ M(\rho) &= \left[F(\rho)' F(\rho)\right]^{-1} F(\rho)'\\ P_N(\rho) &= F_N(\rho) M_N(\rho)\\ P(\rho) &= F(\rho) M(\rho) \end{aligned}$$

For $W = \Sigma^{-1}$ we have that

$$(I_6 - P(1))' W^{1/2} \Sigma W^{1/2} M(1) = 0$$

such that the two linear transformations of $\frac{1}{N}\sum_{i=1}^{N} g(y_i,\theta^*)$ *are orthogonal.*

The probability limit of $\hat{\theta}_2(\rho)$ when the true parameter value equals θ^* is given by

$$\underset{N\to\infty}{\text{plim}}\, \hat{\theta}_2(\rho) = M_2(\rho)\, S(1)\, \theta_2^* \tag{63}$$

where

$$M_2(\rho) = \left[S(\rho)' W S(\rho)\right]^{-1} S(\rho)' W \tag{64}$$

In addition we have that

$$\sqrt{N}\left(\hat{\theta}_2(\rho) - M_2(\rho)\, S(1)\, \theta_2^*\right) \xrightarrow{w} N\left(0, M_2(\rho)\,\Sigma M_2(\rho)'\right) \quad \text{as } N\to\infty \tag{65}$$

This means that $\hat{\theta}_2(1)$ is a $\sqrt{N}$-consistent estimator of θ_2^* since $M_2(1)\, S(1) = I_4$ or more generally that when evaluated at the true value of ρ the estimator $\hat{\theta}_2$ is a $\sqrt{N}$-consistent estimator of the true value of θ_2.

Asymptotic results:
For the first order derivative of $Q_C(\rho)$ we have that

$$N^{3/4} Q_C^{(1)}(1) = 2N^{1/2} g_C(1)' W_N N^{1/4} g_C^{(1)}(1) \xrightarrow{P} 0 \quad \text{as } N\to\infty \tag{66}$$

where we have used that $N^{1/2} g_C(1) \xrightarrow{w} N\left(0,\tilde{\Sigma}\right)$ and that $N^{1/4} g_C^{(1)}(1) \xrightarrow{P} 0$ as $N\to\infty$.

For the second order derivative of $Q_C(\rho)$ we have that

$$\begin{aligned} N^{1/2} Q_C^{(2)}(1)/2! &= N^{1/2} g_C(1)' g_C^{(2)}(1) + N^{1/4} g_C^{(1)}(1)' N^{1/4} g_C^{(1)}(1) \\ &\overset{as}{=} N^{1/2} g_C(1)' g_C^{(2)}(1) \xrightarrow{w} N\left(0, G_2'\tilde{\Sigma} G_2\right) \quad \text{as } N\to\infty \end{aligned} \tag{67}$$

where we have used that $N^{1/2} g_C(1) \xrightarrow{w} N\left(0,\tilde{\Sigma}\right)$, $N^{1/4} g_C^{(1)}(1) \xrightarrow{P} 0$ and that $g_C^{(2)}(1) \xrightarrow{P} G_2$ as $N\to\infty$.

For the third order derivative of $Q_C(\rho)$ we have that

$$N^{1/4} Q_C^{(3)}(1)!/3! = 2/3 N^{1/4} g_C(1)' g_C^{(3)}(1) + N^{1/4} g_C^{(1)}(1)' g_C^{(2)}(1) \xrightarrow{P} 0 \quad \text{as } N\to\infty \tag{68}$$

where we have used that $N^{1/4} g_C(1) \xrightarrow{P} 0$ and $N^{1/4} g_C^{(1)}(1) \xrightarrow{P} 0$ as $N\to\infty$ and also that $g_C^{(2)}(1)$ and $g_C^{(3)}(1)$ both converge in probability as $N\to\infty$.

For the fourth order derivative of $Q_C(\rho)$ we have that

$$\begin{aligned} Q_C^{(4)}(1)/4! &= 1/12 g_C(1)' g_C^{(4)}(1) + 1/3 g_C^{(1)}(1)' g_C^{(3)}(1) + 1/4 g_C^{(2)}(1)' g_C^{(2)}(1) \\ &\overset{as}{=} 1/4 g_C^{(2)}(\rho)' g_C^{(2)}(\rho) \xrightarrow{P} 1/4 G_2' G_2 \quad \text{as } N\to\infty \end{aligned} \tag{69}$$

where we have used that $g_C(1) \xrightarrow{P} 0$ and $g_C^{(1)}(1) \xrightarrow{P} 0$ as $N \to \infty$ and also that $g_C^{(3)}(1)$ and $g_C^{(4)}(1)$ both converge in probability as $N \to \infty$.

For the fifth order derivative of $Q_C(\rho)$ we have that

$$\begin{aligned} & N^{-1/4} Q_C^{(5)}(1)/5! \\ = \; & 2/5! N^{-1/4} g_C(1)' g_C^{(5)}(1) + 10/5! N^{-1/4} g_C^{(1)}(1)' g_C^{(4)}(1) + 20/5! N^{-1/4} g_C^{(2)}(1)' g_C^{(3)}(1) \\ & \xrightarrow{P} 0 \quad \text{as } N \to \infty \end{aligned} \tag{70}$$

where we use arguments similar to the ones above.

This gives the following result

$$\begin{aligned} & N\left(Q_C(\hat{\rho}) - Q_C(1)\right) \\ \overset{as}{=} \; & N^{3/4} Q_C^{(1)}(1) N^{1/4}(\hat{\rho}-1) + N^{1/2} Q_C^{(2)}(1) \left\{N^{1/4}(\hat{\rho}-1)\right\}^2/2! \\ & + N^{1/4} Q_C^{(3)}(1) \left\{N^{1/4}(\hat{\rho}-1)\right\}^3/3! + Q_C^{(4)}(1) \left\{N^{1/4}(\hat{\rho}-1)\right\}^4/4! \\ & + N^{-1/4} Q_C^{(5)}(1) \left\{N^{1/4}(\hat{\rho}-1)\right\}^5/5! \\ \overset{as}{=} \; & N^{1/2} g_C(1)' g_C^{(2)}(1) \left\{N^{1/4}(\hat{\rho}-1)\right\}^2 + 1/4 g_C^{(2)}(1)' g_C^{(2)}(1) \left\{N^{1/4}(\hat{\rho}-1)\right\}^4 \end{aligned} \tag{71}$$

Derivatives of the concentrated moment function:

$$g_C^k(\rho) = \frac{d^k g_C}{d\rho^k}(\rho) \tag{72}$$

The first order derivative

$$g_C^{(1)} = W_N^{1/2}\left(\frac{\partial g}{\partial \rho} + \frac{\partial g}{\partial \theta_2}\hat{\theta}_2^{(1)}\right) = W_N^{1/2}\left(-S^{(1)}(\rho)\hat{\theta}_2(\rho) - S(\rho)\hat{\theta}_2^{(1)}(\rho)\right) \tag{73}$$

The second order derivative

$$g_C^{(2)} = W_N^{1/2}\left(-S^{(2)}(\rho)\hat{\theta}_2(\rho) - 2S^{(1)}(\rho)\hat{\theta}_2^{(1)}(\rho) - S(\rho)\hat{\theta}_2^{(2)}(\rho)\right) \tag{74}$$

The third order derivative

$$g_C^{(3)} = W_N^{1/2}\left(-S^{(3)}(\rho)\hat{\theta}_2(\rho) - 3S^{(2)}(\rho)\hat{\theta}_2^{(1)} - 3S^{(1)}(\rho)\hat{\theta}_2^{(2)}(\rho) - S(\rho)\hat{\theta}_2^{(3)}(\rho)\right) \tag{75}$$

The fourth order derivative

$$g_C^{(4)} = W_N^{1/2}\left(-S^{(4)}(\rho)\hat{\theta}_2(\rho) - 4S^{(3)}(\rho)\hat{\theta}_2^{(1)} - 6S^{(2)}(\rho)\hat{\theta}_2^{(2)} - 3S^{(1)}(\rho)\hat{\theta}_2^{(3)} - S(\rho)\hat{\theta}_2^{(4)}\right) \tag{76}$$

Limiting distribution of $\hat{\theta}_2(\hat{\rho})$:
We use the following definitions

$$F_N(\rho) = W_N^{1/2} S(\rho) \tag{77}$$
$$F_N^{(k)}(\rho) = W_N^{1/2} S^{(k)}(\rho) \quad \text{for } k = 1,2,3,4 \tag{78}$$

The first order derivative of $\hat{\theta}_2(\rho)$:

$$\begin{aligned}
& \hat{\theta}_2^{(1)}(\rho) \\
= \; & \left[F_N(\rho)' F_N(\rho)\right]^{-1} F_N^{(1)}(\rho)' W_N^{1/2} \frac{1}{N} \sum_{i=1}^{N} \text{vech}(y_i y_i') \\
& - \left[F_N(\rho)' F_N(\rho)\right]^{-1} \left[F_N^{(1)}(\rho)' F_N(\rho) + F_N(\rho)' S_N^{(1)}(\rho)\right] \left[F_N(\rho)' F_N(\rho)\right]^{-1} F_N(\rho)' W_N^{1/2} \frac{1}{N} \sum_{i=1}^{N} \text{vech}(y_i y_i') \\
= \; & \left[F_N(\rho)' F_N(\rho)\right]^{-1} \left(F_N^{(1)}(\rho)' W_N^{1/2} \frac{1}{N} \sum_{i=1}^{N} \text{vech}(y_i y_i') - \left[F_N^{(1)}(\rho)' F_N(\rho) + F_N(\rho)' F_N^{(1)}(\rho)\right] \hat{\theta}_2(\rho)\right)
\end{aligned}$$

Using this we have that when $\theta_0 = \theta^*$

$$\begin{aligned}
& N^{1/4} \hat{\theta}_2^{(1)}(1) \\
\overset{as}{=} \; & N^{1/4} \left[F(1)' F(1)\right]^{-1} \left(F^{(1)}(1)' F(1) \theta_2^* - \left[F^{(1)}(1)' F(1) + F(1)' F^{(1)}(1)\right] \theta_2^*\right) \\
= \; & -N^{1/4} \left[F(1)' F(1)\right]^{-1} F(1)' F^{(1)}(1) \theta_2^* \\
= \; & -N^{1/4} q
\end{aligned} \tag{79}$$

where the second line follows by using that $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\text{vech}(y_i y_i') - S(1)\theta_2^*\right) \xrightarrow{w} N(0, \Sigma)$ as $N \to \infty$ and the last line follows from using that $F^{(1)}(1)\theta_2^* = W^{1/2} S^{(1)}(1)\theta_2^* = W^{1/2} S(1) q = F(1) q$ with $q = \left(0, \sigma_\varepsilon^{2*}, \sigma_0^{2*}, -\sigma_\varepsilon^{2*}\right)'$.

The second order derivative of $\hat{\theta}_2(\rho)$:

$$\begin{aligned}
& \hat{\theta}_2^{(2)}(\rho) \\
= \; & \left[F_N(\rho)' F_N(\rho)\right]^{-1} (F_N^{(2)}(\rho)' W_N^{1/2} \frac{1}{N} \sum_{i=1}^{N} \text{vech}(y_i y_i') \\
& - \left[F_N(\rho)' F_N(\rho)\right]^{-1} \left[F_N^{(2)}(\rho)' F_N(\rho) + 2F_N^{(1)}(\rho)' F_N^{(1)}(\rho) + F_N(\rho)' F_N^{(2)}(\rho)\right] \hat{\theta}_2(\rho) \\
& -2 \left[F_N(\rho)' F_N(\rho)\right]^{-1} \left[F_N^{(1)}(\rho)' F_N(\rho) + F_N(\rho)' F_N^{(1)}(\rho)\right] \hat{\theta}_2^{(1)}(\rho)
\end{aligned}$$

Using this we have that when $\theta_0 = \theta^*$

$$\begin{aligned}
& N^{1/4}\hat{\theta}_2^{(2)}(1) \\
\overset{as}{=} \quad & N^{1/4}\left[F(1)' F(1)\right]^{-1} (F^{(2)}(1)' F(1)\theta_2^* - \left[F^{(2)}(1)' F(1) + 2F^{(1)}(1)' F^{(1)}(1) + F(1)' F^{(2)}(1)\right]\theta_2^* \\
& +2\left[F^{(1)}(1)' F(1) + F(1)' F^{(1)}(1)\right] q) \\
= \quad & -\left[F(1)' F(1)\right]^{-1}\left[2F^{(1)}(1)' F^{(1)}(1) + F(1)' F^{(2)}(1)\right] N^{1/4}\theta_2^* \\
& +2\left[F(1)' F(1)\right]^{-1}\left[F^{(1)}(1)' F(1) + F(1)' F^{(1)}(1)\right] N^{1/4} q \\
= \quad & N^{1/4}\left[F(1)' F(1)\right]^{-1} F(1)'\left(2F^{(1)}(1) q - F^{(2)}(1)\theta_2^*\right) \qquad (80)
\end{aligned}$$

where the second line follows by using that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\text{vech}\left(y_i y_i'\right) - S(1)\theta_2^*\right) \overset{w}{\to} N(0,\Sigma)$ as $N \to \infty$ and that $N^{1/4}\left(\hat{\theta}_2(1) - \theta_2^*\right) \overset{P}{\to} 0$ and $N^{1/4}\left(\hat{\theta}_2^{(1)}(1) + q\right) \overset{P}{\to} 0$ as $N \to \infty$ and the third line follows from using that $S(1) q = S^{(1)}(1)\theta_2^*$.

The limiting distribution of $\hat{\theta}_2(\hat{\rho})$ when the true value ρ equals unity and the true value of θ_2 equals $\theta_2^* = \left(\sigma_0^{2*}, 0, 0, \sigma_\varepsilon^{2*}\right)'$

$$\begin{aligned}
& \sqrt{N}\left(\hat{\theta}_2(\hat{\rho}) - \theta_2^* - \hat{\theta}_2^{(1)}(1)(\hat{\rho} - 1) - \hat{\theta}_2^{(2)}(1)(\hat{\rho} - 1)^2/2!\right) \\
= \quad & \sqrt{N}\left(\hat{\theta}_2(\hat{\rho}) - \hat{\theta}_2(1) - \hat{\theta}_2^{(1)}(1)(\hat{\rho} - 1) - \hat{\theta}_2^{(2)}(1)(\hat{\rho} - 1)^2/2! + \hat{\theta}_2(1) - \theta_2^*\right) \\
\overset{as}{=} \quad & \sqrt{N}\left(\hat{\theta}_2(1) - \theta_2^*\right) \\
\overset{w}{\to} \quad & N\left(0, \left[F(1)' F(1)\right]^{-1} F(1)' W^{1/2}\Sigma W^{1/2} F(1)\left[F(1)' F(1)\right]^{-1}\right) \quad \text{as } N \to \infty
\end{aligned}$$

where we have used the result in (65) and that

$$\begin{aligned}
& \sqrt{N}\left(\hat{\theta}_2(\hat{\rho}) - \hat{\theta}_2(1) - \hat{\theta}_2^{(1)}(1)(\hat{\rho} - 1) - \hat{\theta}_2^{(2)}(1)(\hat{\rho} - 1)^2/2!\right) \\
= \quad & N^{-1/4}\hat{\theta}_2^{(3)}(\bar{\rho})/3!\left\{N^{1/4}(\hat{\rho} - 1)\right\}^3 = o_P(1)
\end{aligned}$$

This gives the following asymptotic distribution of $\hat{\theta}_2(\hat{\rho})$

$$\begin{aligned}
N^{1/4}\left(\hat{\theta}_2(\hat{\rho}) - \theta_2^*\right) & \overset{as}{=} \hat{\theta}_2^{(1)}(1) N^{1/4}(\hat{\rho} - 1) + N^{-1/4}\hat{\theta}_2^{(2)}(1)/2!\left\{N^{1/4}(\hat{\rho} - 1)\right\}^2 \\
& \overset{as}{=} -q' N^{1/4}(\hat{\rho} - 1)
\end{aligned}$$

Derivatives of $S(\rho)$:

$$S^{(1)}(\rho) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 2\rho & 0 & 1 & 0 \\ 2\rho & 0 & 2 & 0 \\ 3\rho^2 & 1 & 1+4\rho & 1 \\ 4\rho^3 & 2(1+\rho) & 4\rho+6\rho^2 & 2\rho \end{bmatrix} \quad (81)$$

$$S^{(2)}(\rho) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 6\rho & 0 & 4 & 0 \\ 12\rho^2 & 2 & 4+12\rho & 2 \end{bmatrix} \quad (82)$$

$$S^{(3)}(\rho) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 \\ 24\rho & 0 & 12 & 0 \end{bmatrix} \quad (83)$$

$$S^{(4)}(\rho) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 24 & 0 & 0 & 0 \end{bmatrix} \quad (84)$$

References

Ahn, S.C. and P. Smith, 1995, Efficient estimation of models for dynamic panel data, Journal of Econometrics 68, 5-27.

Ahn, S.C. and P. Smith, 1997, Efficient estimation of models for dynamic panel data: Alternative assumptions and simplified estimation, Journal of Econometrics 76, 309-321.

Andrews, D.W.K., 1999, Estimation when a parameter is on a boundary, Econometrica 67, 1341-1383.

Blundell, R. and S. Bond, 1998, Initial conditions and moment restrictions in a dynamic panel data models, Journal of Econometrics 87, 115-143.

Blundell, R., S. Bond and F. Windmeijer, 2000, Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator, in: Baltagi, B.H. (Ed), Advances in Econometrics vol. 15.

Crepon, B., F. Kramarz and A. Trognon, 1997, Parameters of interest, nuisance parameters and orthogonality conditions: An application to autoregressive error component models, Journal of Econometrics 82, 135-156.

Geyer, C.J., 1994, On the asymptotics of constrained M-estimation, The Annals of Statistics 22, 1993-2010.

Hahn, J., J. Hausman and G. Kursteiner, 2007, Long difference instrumental variables estimation for dynamic panel models with fixed effects, Journal of Econometrics 140, 574-614.

Madsen, E., 2009, GMM estimators and unit root test in the AR(1) panel data model (submitted for publication).

Rotnitzky, A., D.R. Cox, M. Bottai and J. Robins, 2000, Likelihood-based inference with singular information matrix, Bernoulli 6(2), 243-284.

Stock, J.H. and J.H. Wright, 2000, GMM with weak identification, Econometrica 68, 1055-1096.