Journal of Business & Economic Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ubes20

Minimum Distance Estimation of Possibly Noninvertible Moving Average Models

Nikolay Gospodinov[a] & Serena Ng[b]

[a] Research Department, Federal Reserve Bank of Atlanta, Atlanta, GA, 30309

[b] Department of Economics, Columbia University, New York, NY, 10027

Accepted author version posted online: 21 Aug 2014.Published online: 17 Jul 2015.

Minimum Distance Estimation of Possibly Noninvertible Moving Average Models

Nikolay Gospodinov
Research Department, Federal Reserve Bank of Atlanta, Atlanta, GA 30309 (*nikolay.gospodinov@atl.frb.org*)

Serena Ng
Department of Economics, Columbia University, New York, NY 10027 (*sn2294@columbia.edu*)

This article considers estimation of moving average (MA) models with non-Gaussian errors. Information in higher order cumulants allows identification of the parameters without imposing invertibility. By allowing for an unbounded parameter space, the generalized method of moments estimator of the MA(1) model is classical root-T consistent and asymptotically normal when the MA root is inside, outside, and on the unit circle. For more general models where the dependence of the cumulants on the model parameters is analytically intractable, we consider simulation-based estimators with two features. First, in addition to an autoregressive model, new auxiliary regressions that exploit information from the second and higher order moments of the data are considered. Second, the errors used to simulate the model are drawn from a flexible functional form to accommodate a large class of distributions with non-Gaussian features. The proposed simulation estimators are also asymptotically normally distributed without imposing the assumption of invertibility. In the application considered, there is overwhelming evidence of noninvertibility in the Fama-French portfolio returns.

KEY WORDS: Generalized lambda distribution; GMM; Identification; Non-Gaussian errors; Noninvertibility; Simulation-based estimation.

1. INTRODUCTION

Moving average (MA) models can parsimoniously characterize the dynamic behavior of many time series processes. The challenges in estimating MA models are twofold. First, invertible and noninvertible MA processes are observationally equivalent up to the second moments. Second, invertibility restricts all roots of the MA polynomial to be less than or equal to one. This upper bound renders estimators with nonnormal asymptotic distributions when some roots are on or near the unit circle. Existing estimators treat invertible and noninvertible processes separately, requiring the researcher to take a stand on the parameter space of interest. While the estimators are superconsistent under the null hypothesis of an MA unit root, their distributions are not asymptotically pivotal. To our knowledge, no estimator of the MA model exists, which achieves identification without imposing invertibility and yet enables classical inference over the whole parameter space.

Both invertible and noninvertible representations can be consistent with economic theory. For example, if the logarithm of asset price is the sum of a random walk component and a stationary component, the first difference (or asset returns) is generally invertible, but noninvertibility can arise if the variance of the stationary component is large. While noninvertible models are not ruled out by theory, invertibility is often assumed in empirical work because it provides the identification restrictions without which maximum likelihood and covariance structure-based estimation of MA models would not be possible when the data are normally distributed. Invertibility can also be used to narrow the class of equivalent dynamic stochastic general equilibrium (DSGE) models, as in Komunjer and Ng (2011). Obviously, falsely assuming invertibility will yield an inferior fit of the data. It can also lead to spurious estimates of the impulse response coefficients, which are often the objects of interest, as shown by Fernández-Villaverde et al. (2007) using the permanent income model. Hansen and Sargent (1991), Lippi and Reichlin (1993), and Fernández-Villaverde et al. (2007), among others, emphasized the need to verify invertibility because it affects how we interpret what is recovered from the data.

While economic analysis tends to only consider parameter values consistent with invertibility, it is necessary in many science and engineering applications to admit parameter values in the noninvertible range. For example, in analysis of seismic and communication data, noninvertible filters are necessary to recover the earth's reflectivity sequence and to back out the underlying message from a distorted one, respectively. A key finding in these studies is that higher order cumulants are necessary for identification of noninvertible models, implying that the assumption of Gaussian errors must be abandoned. Lii and Rosenblatt (1992) approximated the non-Gaussian likelihood of noninvertible MA models by truncating the representation of the innovations in terms of the observables. Huang and Pawitan (2000) proposed least absolute deviations (LAD) estimation using a Laplace likelihood. This quasi-maximum likelihood (QML) estimator does not require the errors to be Laplace distributed, but they need to have heavy tails. Andrews, Davis, and Breidt (2006, 2007) considered LAD and rank-based estimation of all-pass models, which are special noncausal and/or noninvertible autoregressive and moving average (ARMA)

Journal of Business & Economic Statistics
July 2015, Vol. 33, No. 3
DOI: 10.1080/07350015.2014.955175
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jbes*.

models in which the roots of the autoregressive polynomial are reciprocals of the roots of the MA polynomials. Meitz and Saikkonen (2011) developed maximum likelihood estimation of noninvertible ARMA models with ARCH errors. However, there exist no likelihood-based estimators that have classical properties while admitting an MA unit root in the parameter space.

This article considers estimation of MA models without imposing invertibility. We only require that the errors are non-Gaussian but we do not need to specify the distribution. Identification is achieved by the appropriate use of third and higher order cumulants. In the MA(1) case, "appropriate" means that multiple third moments are necessary, as a single third moment still does not permit identification. In general, identification of possibly noninvertible MA models requires using more unconditional higher order cumulants than the number of parameters in the model. We make use of this identification result to develop generalized method of moments (GMM) estimators that are root-T consistent and asymptotically normal without restricting the MA roots to be strictly inside the unit circle. The estimators minimize the distance between sample-based statistics and their model-based analog. When the model-implied statistics have known functional forms, we have a classical minimum distance estimator.

A drawback of identifying the parameters from the higher order sample moments is that a long span of data is required to precisely estimate the population quantities. This issue is important because for general ARMA(p, q) models, the number of cumulants that needs to be estimated can be quite large. Accordingly, we explore the potential of two simulation estimators in providing bias correction. The first (simulated method of moments, SMM) estimator matches the sample to the simulated unconditional moments as in Duffie and Singleton (1993). The second is a simulated minimum distance (SMD) estimator in the spirit of Gourieroux, Monfort, and Renault (1993) and Gallant and Tauchen (1996). Existing simulation estimators of the MA(1) model impose invertibility and therefore only need the auxiliary parameters from an autoregression to achieve identification. We show that the invertibility assumption can be relaxed but additional auxiliary parameters involving the higher order moments of the data are necessary. In the SMD case, this amounts to estimating an additional auxiliary regression with the second moment of the data as a dependent variable. An important feature of the SMM and SMD estimators is that errors with non-Gaussian features are simulated from the generalized lambda distribution (GLD). These two simulation-based estimators also have classical asymptotic properties regardless of whether the MA roots are inside, outside, or on the unit circle.

The article proceeds as follows. Section 2 highlights two identification problems that arise in MA models. Section 3 presents identification results based on higher order moments of the data. Section 4 discusses GMM estimation of the MA(1) model while Section 5 develops simulation-based estimators for more general MA models. Simulation results and an analysis of the 25 Fama-French portfolio returns are provided in Section 5. Section 6 concludes. Proofs are given in the Appendix.

2. IDENTIFICATION PROBLEMS IN MODELS WITH AN MA COMPONENT

Consider the ARMA (p, q) process:

$$\alpha(L)y_t = \theta(L)e_t, \tag{1}$$

where L is the lag operator such that $L^p y_t = y_{t-p}$ and the lag polynomial $\alpha(L) = 1 - \alpha_1 L - \cdots - \alpha_p L^p$ has no common roots with $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$. Here, y_t can be the error of a regression model

$$\mathbb{Y}_t = x_t'\beta + y_t,$$

where $\mathbb{Y}_t$ is the dependent variable and x_t are exogenous regressors. In the simplest case when $x_t = 1$, y_t is the demeaned data. The process y_t is causal if $\alpha(z) \neq 0$ for all $|z| \leq 1$ on the complex plane. In that case, there exist constants h_j with $\sum_{j=0}^{\infty} |h_j| < \infty$ such that $y_t = \sum_{j=0}^{\infty} h_j e_{t-j}$ for $t = 0, \pm 1, \ldots$ Thus, all MA models are causal. The process is invertible if $\theta(z) \neq 0$ for all $|z| \leq 1$; see Brockwell and Davies (1991). In control theory and the engineering literature, an invertible process is said to have minimum phase.

Our interest is in estimating MA models without prior knowledge about invertibility. The distinction between invertible and noninvertible processes is best illustrated by considering the MA(1) model defined by

$$y_t = e_t + \theta e_{t-1}, \tag{2}$$

where $e_t = \sigma\varepsilon_t$ and $\varepsilon_t \sim \text{iid}(0, 1)$ with $\kappa_3 = E(\varepsilon_t^3)$ and $\kappa_4 = E(\varepsilon_t^4)$. The invertibility condition is satisfied if $|\theta| < 1$. In that case, the inverse of $\theta(L)$ has a convergent series expansion in positive powers of the lag operator L. Then, we can express y_t as $\pi(L)y_t = e_t$ with $\pi(L) = \sum_{j=0}^{\infty}(-\theta L)^j$. This infinite autoregressive representation of y_t implies that the span of e_t and its history coincide with that of y_t, which is observed by the econometrician. When $|\theta| > 1$, the inverse polynomial is $\sum_{j=0}^{\infty}(-\theta L)^{-j-1}$, implying that y_t is a function of future values of y_t, which is not useful for forecasting. This argument is often used to justify the assumption of invertibility. It is, however, misleading to classify invertible processes according to the value of θ alone. Consider another MA(1) process y_t represented by

$$y_t = \theta e_t + e_{t-1}. \tag{3}$$

Even if θ in (3) is less than one, the inverse of $\theta(L) = (\theta + L)$ is not convergent. Furthermore, the errors from a projection of y_t on lags of y_t have different time series properties depending on whether the data are generated by (2) or (3).

Identification and estimation of models with an MA component are difficult because of two problems that are best understood by focusing on the MA(1) case. The first identification problem concerns θ at or near unity. When the MA parameter θ is near the unit circle, the Gaussian maximum likelihood (ML) estimator takes values exactly on the boundary of the invertibility region with positive probability in finite samples. This point probability mass at unity (the so-called "pile-up" problem) arises from the symmetry of the likelihood function around one and the small sample deficiency to identify all the critical points of the likelihood function in the vicinity of the noninvertibility boundary; see Sargan and Bhargava (1983), Anderson and Takemura (1986), Davis and Dunsmuir (1996), Gospodinov (2002), and Davis and Song (2011).

The second identification problem arises because covariance stationary processes are completely characterized by the first and second moments of the observables. The Gaussian likelihood for an MA(1) model with $\mathcal{L}(\theta, \sigma^2)$ is the same as one with $\mathcal{L}(1/\theta, \theta^2\sigma^2)$. The observational equivalence of the covariance structure of invertible and noninvertible processes also implies that the projection coefficients in $\pi(L)$ are the same regardless of whether θ is less than or greater than one. Thus, θ cannot be recovered from the coefficients of $\pi(L)$ without additional assumptions.

This observational equivalence problem can be further elicited from a frequency domain perspective. If we take as a starting point $y_t = h(L)e_t = \sum_{j=-\infty}^{\infty} h_j e_{t-j}$, the frequency response function of the filter is

$$H(\omega) = \sum h_j \exp(-i\omega j) = |H(\omega)| \exp^{-i\delta(\omega)},$$

where $|H(\omega)|$ is the amplitude and $\delta(\omega)$ is the phase response of the filter. For ARMA models, $h(z) = \frac{\theta(z)}{\alpha(z)} = \sum_{j=-\infty}^{\infty} h_j z^j$. The amplitude is usually constant for given ω and tends toward zero outside the interval $[0, \pi]$. For given $a > 0$, the phase δ_0 is indistinguishable from $\delta(\omega) = \delta_0 + a\omega$ for any $\omega \in [0, \pi]$. Recovering e_t from the second-order spectrum

$$S_2(z) = \sigma^2 |H(z)|^2$$

is problematic because $S_2(z)$ is proportional to the amplitude $|H(z)|^2$ with no information about the phase $\delta(\omega)$. The second-order spectrum is thus said to be phase-blind. As argued by Lii and Rosenblatt (1982), one can flip the roots of $\alpha(z)$ and $\theta(z)$ without affecting the modulus of the transfer function. With real distinct roots, there are 2^{p+q} ways of specifying the roots without changing the probability structure of y_t.

3. CUMULANT-BASED IDENTIFICATION OF NONINVERTIBLE MODELS

Econometric analysis on identification largely follows the pioneering work of Fisher (1961, 1965) and Rothenberg (1971) in fully parametric/likelihood settings. These authors recast the identification problem as one of finding a unique solution to a system of nonlinear equations. For nonlinear models, a sufficient condition is that the Jacobian matrix of the first partial derivatives is of full column rank. See Dufour and Hsiao (2008) for a survey. However, local identification is still possible if the rank condition fails by exploiting restrictions on the higher order derivatives, as shown in Sargan (1983) and Dovonon and Renault (2013). To obtain results for global identification, Rothenberg (1971, Theorem 7) imposed additional conditions to ensure that the optimization problem is well behaved. In a semiparametric setting when the distribution of the errors is not specified, identification results are limited, but the rank of the derivative matrix remains to be a sufficient condition for local identification (Newey and McFadden 1994). Komunjer (2012) showed that global identification from moment restrictions is possible even when the derivative matrix has a deficient rank, provided that this happens only over sufficiently small regions in the parameter space.

More precisely, let $\gamma \in \Gamma$ be a $K \times 1$ parameter vector of interest, where the parameter space Γ is a compact subset of the K-dimensional Euclidean space $\mathcal{R}^K$. In the case of an ARMA(p, q) model defined by (1), $\gamma = (\alpha_1, \ldots, \alpha_p, \theta_1, \ldots, \theta_q, \sigma^2)'$. Let γ_0 be the true value of γ and $g(\gamma) \in \mathcal{G} \subset \mathcal{R}^L$ denote L $(L \geq K)$ moments, which can be used to infer the value of γ_0. Identification hinges on a well-behaved mapping from the space of γ to the space of moment conditions $g(\cdot)$.

Definition 1. Let $g(\gamma) : \mathcal{R}^K \to \mathcal{R}^L$ be a mapping from γ to $g(\gamma)$ and let $G(\gamma) = \partial g(\gamma)/\partial\gamma'$ with $G_0 \equiv G(\gamma_0)$. Then, γ_0 is globally identified from $g(\gamma)$ if $g(\cdot)$ is injective and is locally identified if the matrix of partial derivatives G_0 has full column rank.

From Definition 1, γ_1 and γ_2 are observationally equivalent if $g(\gamma_1) = g(\gamma_2)$, that is, $g(\cdot)$ is not injective. Section 3.1 shows in the context of an MA(1) model that second moments cannot be used to define a vector $g(\gamma)$ that identifies γ without imposing invertibility. However, possibly noninvertible models can be identified if $g(\gamma)$ is allowed to include higher order moments/cumulants. Sections 3.2 and 3.3 generalize the results to MA(q) and ARMA(p, q) models.

3.1 MA(1) Model

This subsection provides a traditional identification analysis of the zero mean MA(1) model. Let $\gamma = (\theta, \sigma^2)'$. The data y_t are a function of the true value γ_0. For the MA(1) model, $E(y_t y_{t-1}) = 0$ for $j \geq 2$. Consider the population identification problem using only second moments of y_t:

$$g_2(\gamma) = \begin{pmatrix} g_{21} \\ g_{22} \end{pmatrix} = \begin{pmatrix} E(y_t y_{t-1}) \\ E(y_t^2) \end{pmatrix} - \begin{pmatrix} \theta\sigma^2 \\ (1+\theta^2)\sigma^2 \end{pmatrix}.$$

The moment vector $g_2(\gamma)$ is the difference between the population second moments and the moments implied by the MA(1) model. If the assumption that the data are generated by the MA(1) model is correct, $g_2(\gamma)$ evaluated at the true value of γ is zero: $g_2(\gamma_0) = 0$. Under Gaussianity of the errors, these moments fully characterize the covariance structure of y_t. However, $g_2(\gamma)$ assumes the same value for $\gamma_1 = (\theta, \sigma^2)'$ and $\gamma_2 = (1/\theta, \theta^2\sigma^2)'$. For example, if $\gamma_1 = (\theta = 0.5, \sigma^2 = 1)'$ and $\gamma_2 = (\theta = 2, \sigma^2 = 0.25)'$, $g_2(\gamma_1) = g_2(\gamma_2)$. Parameters that are not identifiable from the population moments are not consistently estimable.

The problem that the mapping $g_2(\cdot)$ is not injective is typically handled by imposing invertibility, thereby restricting the parameter space to $\Gamma^R = [-1, 1] \times [\sigma_L^2, \sigma_H^2]$. But there is still a problem because the derivative matrix of $g(\gamma)$ with respect to γ is not full rank everywhere in Γ^R. The determinant of

$$G(\gamma) = \begin{pmatrix} \sigma^2 & \theta \\ 2\theta\sigma^2 & (1+\theta^2) \end{pmatrix} \tag{4}$$

is zero when $|\theta| = 1$. This is responsible for the pile-up problem discussed earlier. Furthermore, $|\theta| = 1$ lies on the boundary of the parameter space. As a consequence, the Gaussian maximum likelihood estimator and estimators based on second moments are not uniformly asymptotically normal; see Davis and Dunsmuir (1996). Note, however, that the two problems with the MA(1) model, namely, inconsistency due to nonidentification and nonnormality due to a unit root, do not arise if there is prior knowledge about σ^2. We will revisit this observation in Section 4.1.

While the second moments of the data do not identify $\gamma = (\theta, \sigma^2)'$, would the three nonzero third moments given by

$$g_3(\gamma) = \begin{pmatrix} g_{31} \\ g_{32} \\ g_{33} \end{pmatrix} = \begin{pmatrix} E(y_t^3) \\ E(y_t^2 y_{t-1}) \\ E(y_t y_{t-1}^2) \end{pmatrix} - \begin{pmatrix} (1+\theta^3)\sigma^3\kappa_3, \\ \theta^2\sigma^3\kappa_3 \\ \theta\sigma^3\kappa_3 \end{pmatrix}$$

achieve identification? The following lemma provides an answer to this question.

Lemma 1. Consider the MA(1) model $y_t = e_t + \theta e_{t-1}$ with $e_t = \sigma\varepsilon_t$. Suppose that $\varepsilon_t \sim \text{iid}(0, 1)$ with $\kappa_3 = E(\varepsilon_t^3)$. Assume that $\theta \neq 0$, $\kappa_3 \neq 0$, and $E|\varepsilon_t|^3 < \infty$. Then,

(a) $g(\gamma) = (g_2', g_{32})'$ is not injective for any $\gamma = (\theta, \sigma^2, \kappa_3)' \in \Gamma$.
(b) $g(\gamma) = (g_2', g_{3j})'$ for $j = 1, 2,$ or 3 cannot locally identify γ when $|\theta| = 1$ for any σ^2 and κ_3.

In Lemma 1, $g_3(\cdot)$ and $\gamma = (\theta, \sigma^2, \kappa_3)'$ are of the same dimension. Part (a) states that there always exist $\gamma_1, \gamma_2 \in \Gamma$ that are observationally equivalent in the sense that they generate the same moments. For example, $\gamma_1 = (\theta, \sigma^2, \kappa_3)'$ and $\gamma_2 = (1/\theta, \theta^2\sigma^2, \theta\kappa_3)'$ both imply the same $(E(y_t y_{t-1}), E(y_t^2), E(y_t^2 y_{t-1}))'$. Part (b) of Lemma 1 follows from the fact that the determinant of the derivative matrix is zero at $|\theta| = 1$. As a result, a single third moment cannot be guaranteed to identify both κ_3 and the parameters of the MA(1) model θ and σ^2. Global and local identification of θ at $|\theta| = 1$ requires use of information in the remaining two third-order moments. In particular, the derivative matrix of $g(\gamma) = (g_2', g_3')'$ with respect to $\gamma = (\theta, \sigma^2, \kappa_3)'$ is of full column rank everywhere in Γ including $|\theta| = 1$. However, since $g(\cdot)$ is of dimension five, this together with Lemma 1 implies that γ can only be overidentified if $\kappa_3 \neq 0$. The next subsection describes a general procedure, based on higher order cumulants, for identifying the parameters of MA(q) and ARMA (p, q) models.

3.2 The MA(q) Model

The insight from the MA(1) analysis that the parameters of the model cannot be exactly identified but can be over-identified with an appropriate choice of higher order moments extends to MA(q) models. But for MA(q) models, the moments of the process are nonlinear functions of the model parameters and verifying global and local identification is more challenging. Our analysis is built on results from the statistical engineering literature.

Let $c_\ell(\tau_1, \tau_2, \ldots, \tau_{\ell-1})$ be the ℓth ($\ell \geq 2$)-order cumulant of a zero-mean stationary and ergodic process y_t. The second- and third-order cumulants of y_t are given by

$$c_2(\tau_1) = E(y_t y_{t+\tau_1}),$$
$$c_3(\tau_1, \tau_2) = E(y_t y_{t+\tau_1} y_{t+\tau_2}).$$

If $y_t = h(L)e_t$ and $e_t = \sigma\varepsilon_t$ is a mean-zero iid process, we have

$$c_\ell(\tau_1, \ldots, \tau_{\ell-1}) = \eta_\ell \sum_{i=0}^{\infty} h_i h_{i+\tau_1} \ldots h_{i+\tau_{\ell-1}}, \tag{5}$$

where $\eta_\ell = c_{\ell e}(0, 0, \ldots, 0)$ denotes the ℓth-order cumulant of e_t with $\eta_2 = \sigma^2$, $\eta_3 = \kappa_3\sigma^3$, and $\eta_4 = \sigma^4(\kappa_4 - 3)$. Thus, the cumulants η_ℓ ($\ell \geq 3$) measure the distance of e_t (and hence of y_t) from Gaussianity. If $\tau_1 = \tau_2 = \cdots = \tau_\ell = \tau$, then $c_\ell(\tau) = c_\ell(\tau, \ldots, \tau)$ is known as the diagonal slice of the ℓth-order cumulant of y_t.

Higher order cumulants are useful for identification of possibly noninvertible models because the Fourier transform of $c_\ell(\tau_1, \tau_2, \ldots, \tau_{\ell-1})$ is the ℓth-order polyspectrum

$$S_\ell(\omega_1, \ldots, \omega_{\ell-1}) = \eta_\ell H(\omega_1) \ldots H(\omega_{\ell-1}) H\left(-\sum_{i=1}^{\ell-1} \omega_i\right). \tag{6}$$

Recovery of phase information necessarily requires that e_t has non-Gaussian features. In other words, η_ℓ must exist and is nonzero for some $\ell \geq 3$ for recovery of the phase function; see Lii and Rosenblatt (1982, Lemma 1), Giannakis and Swami (1992), Giannakis and Mendel (1989), Mendel (1991), Tugnait (1986), and Ramsey and Montenegro (1992).

To establish that the MA(q) parameters are identifiable from cumulants of a particular order ℓ, the typical starting point is to generate identities that link the second and higher order cumulants to the parameters of the model. Different identities exist for different choice of $\tau_1, \tau_2, \ldots, \tau_{\ell-1}$. Mendel (1991) provided a survey of the methods used in the engineering literature. One of the simplest and earliest ideas is to consider the diagonal slice of the third-order cumulants, which implies the following relation between the population cumulants and the $q+1$ vector of parameters $\gamma = (\theta_1, \ldots, \theta_q, \kappa_3\sigma)'$:

$$\sum_{j=1}^{q} \theta_j c_3(\tau - j) - \kappa_3\sigma \sum_{j=0}^{q} \theta_j^2 c_2(\tau - j) + c_3(\tau) = 0,$$
$$-q \leq \tau \leq 2q. \tag{7}$$

Define

$$\beta(\gamma) = \left(\theta_1, \ldots, \theta_q, \kappa_3\sigma, \kappa_3\sigma\theta_1^2, \ldots, \kappa_3\sigma\theta_q^2\right)',$$

and

$$b = [-c_3(-q) \quad -c_3(-q+1) \quad \cdots \quad -c_3(0) \quad -c_3(1) \quad \cdots \quad -c_3(q-1) \quad -c_3(q) \quad 0 \quad 0 \quad \cdots \quad 0]'.$$

The system of Equation (7) can be expressed as

$$A\beta(\gamma) = b. \tag{8}$$

The reason why (8) is useful for identification is that $A\beta(\gamma) = b$ is an over-identified system of $3q+1$ equations in $2q+1$ unknowns $\beta(\gamma)$. The parameters γ are identifiable if $\beta(\gamma)$ can be solved from (8). Given that the derivative matrix of $\beta(\gamma)$ with respect to γ has rank $q+1$, the identification problem reduces to the verification of the column rank of the matrix A (given in the Appendix).

Lemma 2. Consider the MA(q) process $y_t = e_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}$, where $e_t = \sigma\varepsilon_t$, $\varepsilon_t \sim \text{iid}(0, 1)$ with

$\kappa_3 = E(\varepsilon_t^3) \neq 0$ and $E|\varepsilon_t|^3 < \infty$. Let $c_\ell(\tau)$ denote the diagonal slice of the ℓth-order cumulant of y_t. If $c_2(q)$ and $c_3(q)$ are nonzero, then the matrix A has full column rank $2q + 1$.

A proof is given in the Appendix. Full rank of the matrix A enables identification of $\beta(\gamma)$ and subsequently of γ. This requires that the qth autocorrelation $c_2(q)$ is nonzero, and also that $c_3(q) \neq 0$. In view of the definition of $c_3(q)$ in (5), it is clear that skewness in e_t is necessary for identification of γ. A similar idea can be used to analyze identification using fourth-order cumulants, defined as $c_4(\tau_1, \tau_2, \tau_3) = E(y_t y_{t+\tau_1} y_{t+\tau_2} y_{t+\tau_3}) - c_2(\tau_1)c_2(\tau_2 - \tau_3) - c_2(\tau_2)c_2(\tau_3 - \tau_1) - c_2(\tau_3)c_2(\tau_1 - \tau_2)$. For example, the diagonal slice of the fourth-order cumulants yields a system of equations given by

$$\sum_{i=1}^{q} \theta_i c_4(\tau - i) - \sigma^2(\kappa_4 - 3) \sum_{i=0}^{q} \theta_i^3 c_2(\tau - i) = -c_4(\tau). \quad (9)$$

As shown in the Appendix, $\beta(\theta_1, \ldots, \theta_q, \sigma^2(\kappa_4 - 3)) = (\theta_1, \ldots, \theta_q, \sigma^2(\kappa_4 - 3), \sigma^2(\kappa_4 - 3)\theta_1^3, \ldots, \sigma^2(\kappa_4 - 3)\theta_q^3)'$ is identifiable provided that $(\kappa_4 - 3)$, $c_2(q)$, and $c_4(q)$ are nonzero. Nondiagonal slices of the fourth-order cumulants were considered by Friedlander and Porat (1990) and Na et al. (1995).

Giannakis and Mendel (1989, p. 364) made use of the structure of the A matrix to recursively compute the parameters in $\beta(\gamma)$ and hence γ. This algorithm treats $\kappa_3\sigma\theta_1^2, \ldots, \kappa_3\sigma\theta_q^2$ as free parameters when, in fact, they are not. Although the method is not efficient or practical for estimation, the approach is one of the first to suggest the possibility of identification of MA models using cumulants. Tugnait (1995) subsequently obtained closed-form expressions for the MA parameters using $c_3(\tau, \tau + q)$ and autocovariances. Friedlander and Porat (1990) proposed an optimal minimum distance estimation of the system (8) although this method still cannot separately identify the parameters κ_3 (or κ_4) and σ^2. As we will see below, this approach is a restricted version of our proposed GMM method.

3.3 ARMA(*p*, *q*) Model

The previous two subsections have focused on MA(q) models because the p parameters in the autoregressive polynomial $\alpha(L)$ can be easily identified. Consider the ARMA(p, q) model

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} = e_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q},$$

where $e_t \sim \text{iid}(0, \sigma^2)$. If e_t were Gaussian, one can exploit the fact that $Ee_{t-q}y_{t-j} = 0$ for $j > q$, or equivalently, $c_2(\tau + 1) - \sum_{k=0}^{p-1} c_2(\tau - k)\alpha_{k+1} = 0$ for $\tau \in [q, q + p]$. This leads to the system of equations

$$\begin{pmatrix} c_2(q+1) \\ c_2(q+2) \\ \vdots \\ c_2(q+p) \end{pmatrix} = \begin{pmatrix} c_2(q) & c_2(q+1) & \ldots & c_2(q+p-1) \\ c_2(q+1) & c_2(q) & \ldots & \\ \vdots & & & \vdots \\ c_2(q+p-1) & \ldots & & c_2(q) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix}. \quad (10)$$

The $p \times p$ Toeplitz matrix on the right-hand side is a submatrix of autocovariances and hence full rank. Thus, α is identifiable. By considering the spectrum at p frequencies, identities can also be derived in the frequency domain. If e_t were non-Gaussian, the AR coefficients of an ARMA process can still be uniquely determined from the equations $\sum_{i=0}^{p}\sum_{j=0}^{p} \alpha_i\alpha_j c_\ell(\tau_1 - i, \tau_2 - j, \tau_3, \ldots, \tau_{\ell-1}) = 0$ for $\ell \geq 3$ and $|\tau_1 - \tau_2| > q$. The idea of using cumulants to identify the autoregressive parameters seems to date back to Akaike (1966), see Mendel (1991, p. 281) and Theorem 2 of Giannakis and Swami (1992).

The question then arises as to whether $(\alpha_1, \ldots, \alpha_p, \theta_1, \ldots, \theta_q)'$ can be jointly identified from the third-order cumulants alone. The $A\beta(\gamma) = b$ framework presented above requires that q is finite and hence does not work for ARMA(p, q) models. Assuming that the ARMA model has no common factors (hence, it is irreducible), the following lemma, adapted from Tugnait (1995), provides sufficient conditions for identifiability of the parameters of ARMA(p, q) models.

Lemma 3. Assume that the ARMA(p, q) process $(1 - \alpha_1 L - \cdots - \alpha_p L^p)y_t = (1 + \theta_1 L + \cdots + \theta_q L^q)e_t$ is irreducible and satisfies $\sum_{i=0}^{p} \alpha_i z^i \neq 0$ for $|z| = 1$, where $e_t = \sigma\varepsilon_t$, $\varepsilon_t \sim \text{iid}(0, 1)$ with $\kappa_3 = E(\varepsilon_t^3) \neq 0$ and $E|\varepsilon_t|^3 < \infty$. Let $c_\ell(\tau)$ denote the diagonal slice of the ℓth-order cumulant of the MA($p+q$) process $(1 - \alpha_1 L - \cdots - \alpha_p L^p)(1 + \theta_1 L + \cdots + \theta_q L^q)e_t$ and assume that $c_2(p+q)$ and $c_3(p+q)$ are nonzero. Then, the parameter vector $(\alpha_1, \ldots, \alpha_p, \theta_1, \ldots, \theta_q)'$ of the ARMA(p, q) process is identifiable from the second and third cumulants of the MA($p+q$) process.

The thrust of the argument, elaborated in the Appendix, is that observational equivalence of the two ARMA(p, q) process amounts to equivalence of two appropriately defined MA($p+q$) processes, say, z, parameterized by $p+q$ vectors Θ_1 and Θ_2, respectively. But from Tugnait (1995), two MA($p+q$) processes are equivalent if $c_{3z}(\tau_1, p+q|\Theta_1) = c_{3z}(\tau_1, p+q|\Theta_2)$ for $0 \leq \tau_1 \leq p+q$. We can now exploit results from the previous subsection. Tugnait (1995) used information in the nondiagonal slices to isolate the smallest number of third and higher order cumulants that are sufficient for identification of ARMA parameters.

The representation $A\beta(\gamma) = b$ provides a transparent way to see how higher order cumulants can be used to recover the parameters of the model without imposing invertibility. However, this approach may use more cumulants than is necessary. To see why, (5) implies that for an MA(q) process, $c_3(q, k) = \kappa_3\sigma^3\theta_q\theta_k$ and $c_3(q, 0) = \kappa_3\sigma^3\theta_q$. It immediately follows that $\theta_k = \frac{c_3(q,k)}{c_3(q,0)}$. This so-called $C(q, k)$ formula suggests that only $q + 1$ third-order cumulants $c_3(q, \tau)$ for $0 \leq \tau \leq q$ are necessary and sufficient for identification of $\theta_1, \ldots, \theta_q$ if $\kappa_3 \neq 0$, which is smaller than the number of equations in the $A\beta(\gamma) = b$ system.

The key point in this section to highlight is that once non-Gaussian features are allowed, identification of noninvertible models is possible from the higher order cumulants of the data. In practice, we would want to use the covariance structure along with identities based on the third- and fourth-order cumulants. Using information in the third or fourth cumulants alone would be inefficient, even though identification is possible. This is because the covariance structure would have been sufficient for identification if invertibility was imposed, and the fourth-order cumulants can be useful when the error distribution is (near-) symmetric. The identities considered shed light on which order cumulants are required for identification. For example, in the MA(1) case, the $A\beta(\gamma) = b$ system

$$\begin{bmatrix} 0 & -c_2(1) & 0 \\ c_3(-1) & -c_2(0) & -c_2(1) \\ c_3(0) & -c_2(1) & -c_2(0) \\ c_3(1) & 0 & -c_2(1) \end{bmatrix} \begin{bmatrix} \theta \\ \kappa_3\sigma \\ \kappa_3\sigma\theta^2 \end{bmatrix} = \begin{bmatrix} -c_3(-1) \\ -c_3(0) \\ -c_3(1) \\ 0 \end{bmatrix} \tag{11}$$

tells us that the third-order cumulants $c_3(1), c_3(0), c_3(-1)$ will be needed to identify the MA(1) parameters. This is used to guide estimation, which is the subject of the next section.

4. GMM ESTIMATION

The results in Section 3 suggest to estimate the parameters of ARMA(p, q) models by matching second and higher order cumulants. Friedlander and Porat (1990, p. 30) proposed a two-step procedure for estimating ARMA(p, q) models where the AR parameters are obtained first from the autocovariances (spectrum) of the process and the MA parameters are then estimated from the filtered process using information in the higher order cumulants (a similar estimation strategy has been proposed by an anonymous referee). Our proposed estimation strategy is similar in spirit but it estimates all of the unknown parameters in one step.

Let $g_t(\gamma)$ be conditions characterizing the model parameterized by γ and such that at the true value γ_0, $E[g_t(\gamma_0)] = 0$. Given data $\mathbf{y} \equiv (y_1, \ldots, y_T)'$, one can construct $\overline{g}(\gamma) = \frac{1}{T}\sum_{t=1}^{T} g_t(\gamma)$, the sample analog of $g(\gamma) = E[g_t(\gamma)]$. Let $\widehat{\Omega}$ denote a consistent estimate of the positive definite matrix $\Omega = \lim_{T\to\infty} \mathrm{var}(\sqrt{T}\,\overline{g}(\gamma_0))$. The optimal GMM estimator of γ is defined as

$$\widehat{\gamma} = \arg\min_{\gamma} \overline{g}(\gamma)'\widehat{\Omega}^{-1}\overline{g}(\gamma). \tag{12}$$

Full rank of the derivative matrix $G(\gamma) = \frac{\partial g(\gamma)}{\partial \gamma'}$ evaluated at γ_0 is sufficient for γ_0 to be a unique solution to the system of nonlinear equations characterized by $G(\gamma)'\Omega^{-1}g(\gamma) = 0$. The full rank condition in the neighborhood of γ_0 is also necessary for the estimator to be asymptotically normal. Under the assumptions given by Newey and McFadden (1994),

$$\sqrt{T}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} N\big(0, (G(\gamma_0)'\Omega^{-1}G(\gamma_0))^{-1}\big). \tag{13}$$

Consistent estimation of possibly noninvertible ARMA(p, q) models depends on the choice $g_t(\gamma)$. We consider three possibilities beginning with a classical GMM estimator.

For the MA(1) model, let $\gamma = (\theta, \sigma^2, \kappa_3)'$ be the parameters to be estimated and define

$$\overline{g}(\gamma) = \overline{m}(\gamma_0) - m(\gamma),$$

where $\overline{m}(\gamma_0) = \frac{1}{T}\sum_{t=1}^{T} m_t(\gamma_0)$ is a consistent estimate of $E[m_t(\gamma_0)]$. The identification results in Lemma 1 and (14) suggest to consider

$$E[m_t(\gamma_0)] = \begin{pmatrix} E(y_t y_{t-1}) \\ E(y_t^2) \\ E(y_t^2 y_{t-1}) \\ E(y_t^3) \\ E(y_t y_{t-1}^2) \end{pmatrix} = \begin{pmatrix} c_2(1) \\ c_2(0) \\ c_3(1) \\ c_3(0) \\ c_3(-1) \end{pmatrix},$$

$$m(\gamma) = \begin{pmatrix} \theta\sigma^2 \\ (1+\theta^2)\sigma^2 \\ \theta^2\sigma^3\kappa_3 \\ (1+\theta^3)\sigma^3\kappa_3 \\ \theta\sigma^3\kappa_3 \end{pmatrix}. \tag{14}$$

Note that the equations in (11) are particular linear combinations of the moment conditions in (14). The conditions in Lemma 2 that $c_2(1) \neq 0$ and $c_3(1) \neq 0$ correspond to the conditions $\theta \neq 0$ and $\kappa_3 \neq 0$ in Lemma 1.

Proposition 1. Consider the MA(1) model. Suppose that in addition to the assumptions in Lemma 1, we have that $E|e_t|^6 < \infty$ and γ_0 is in the interior of the compact parameter space Γ. Also, assume that $\sqrt{T}\,\overline{g}(\gamma_0) \xrightarrow{d} N(0, \Omega)$ and $\widehat{G}(\gamma) = \partial\overline{g}(\gamma)/\partial\gamma'$ converges uniformly to $G(\gamma)$ over $\gamma \in \Gamma$. Then, $\widehat{\gamma}$ is $\sqrt{T}$ consistent with asymptotic distribution given by (13).

The derivative matrix $G(\gamma) = \frac{\partial g(\gamma)}{\partial \gamma'}$ is of full column rank everywhere in Γ (even at $|\theta| = 1$). As a result, this GMM estimator is root-T consistent and asymptotically normal.

4.1 Finite-Sample Properties of the GMM Estimator

To illustrate the finite-sample properties of the GMM estimator, data with $T = 500$ observations are generated from an MA(1) model $y_t = e_t + \theta e_{t-1}$ and $e_t = \sigma\varepsilon_t$, where ε_t is iid(0, 1) and follows a GLD, which will be further discussed in Section 5.1. For now, it suffices to note that GLD distributions can be characterized by a skewness parameter κ_3 and a kurtosis parameter κ_4. The true values of the parameters are $\theta = 0.5, 0.7, 1, 1.5$, and 2, $\sigma = 1$, $\kappa_3 = 0, 0.35, 0.6$, and 0.85, and $\kappa_4 = 3$. The results are invariant to the choice of σ. Lack of identification of γ arises when $\kappa_3 = 0$ and weak to intermediate identification occurs when $\kappa_3 = 0.35$ and 0.6. Unreported numerical results revealed that the estimator based on the moment conditions (14) possesses substantially better finite-sample properties than the estimator based on (11). We only consider the finite-sample properties of the estimator for the MA(1) model when the orthogonality conditions are both necessary and sufficient for identification.

Table 1 presents the mean, the median, and the standard deviation of three estimators of θ over 1000 Monte Carlo replications. The first is the GMM estimator of $\gamma = (\theta, \sigma^2, \kappa_3)'$, which uses (14) as moment conditions. The second is the infeasible GMM estimator based on (14) but assumes σ^2 is known and estimates only $(\theta, \kappa_3)'$. As discussed earlier, fixing σ^2 solves the identification problem in the MA(1) model, and by not imposing invertibility, $|\theta| = 1$ is not on the boundary of the parameter space for γ. We will demonstrate

Table 1. GMM and Gaussian QML estimates of θ from MA(1) model with possibly asymmetric errors

	GMM estimator				Gaussian QML estimator				Infeasible GMM estimator			
θ_0	Mean	Med.	$P(\widehat{\theta} \geq 1)$	Std.	Mean	Med.	$P(\widehat{\theta} \geq 1)$	Std.	Mean	Med.	$P(\widehat{\theta} \geq 1)$	Std.
						$\kappa_3 = 0$						
0.5	1.392	1.692	0.578	0.790	0.500	0.502	0.000	0.040	0.489	0.486	0.000	0.071
0.7	1.152	1.117	0.564	0.428	0.700	0.701	0.000	0.033	0.674	0.675	0.000	0.084
1.0	1.057	1.004	0.509	0.279	0.965	0.971	0.063	0.028	0.970	0.974	0.386	0.082
1.5	1.144	1.105	0.547	0.467	0.666	0.667	0.000	0.034	1.473	1.471	1.000	0.073
2.0	1.353	1.600	0.563	0.783	0.500	0.501	0.000	0.040	1.969	1.967	1.000	0.081
						$\kappa_3 = 0.35$						
0.5	0.823	0.518	0.223	0.615	0.500	0.500	0.000	0.040	0.488	0.484	0.000	0.071
0.7	0.903	0.773	0.262	0.368	0.699	0.700	0.000	0.033	0.675	0.673	0.000	0.085
1.0	1.057	1.020	0.543	0.264	0.964	0.969	0.053	0.028	0.972	0.976	0.377	0.081
1.5	1.367	1.427	0.808	0.414	0.666	0.667	0.000	0.034	1.475	1.474	1.000	0.073
2.0	1.757	1.950	0.827	0.642	0.500	0.501	0.000	0.040	1.971	1.969	1.000	0.080
						$\kappa_3 = 0.6$						
0.5	0.552	0.493	0.034	0.260	0.500	0.501	0.000	0.040	0.488	0.485	0.000	0.071
0.7	0.738	0.690	0.062	0.203	0.699	0.700	0.000	0.033	0.677	0.673	0.000	0.085
1.0	1.042	1.009	0.528	0.237	0.964	0.968	0.048	0.028	0.975	0.982	0.389	0.077
1.5	1.514	1.527	0.964	0.307	0.666	0.667	0.000	0.034	1.478	1.478	1.000	0.069
2.0	1.986	2.039	0.969	0.423	0.500	0.501	0.000	0.040	1.975	1.973	1.000	0.076
						$\kappa_3 = 0.85$						
0.5	0.511	0.487	0.003	0.121	0.500	0.500	0.000	0.040	0.489	0.485	0.000	0.069
0.7	0.688	0.674	0.003	0.118	0.699	0.699	0.000	0.033	0.678	0.677	0.000	0.084
1.0	1.012	0.999	0.496	0.187	0.964	0.966	0.046	0.027	0.978	0.987	0.416	0.072
1.5	1.556	1.544	0.997	0.268	0.666	0.667	0.000	0.034	1.482	1.483	1.000	0.063
2.0	2.025	2.043	0.993	0.366	0.500	0.501	0.000	0.040	1.980	1.979	1.000	0.070

NOTES: The table reports the mean, median (med.), probability that $\widehat{\theta} \geq 1$, and standard deviation (std.) of the GMM, Gaussian quasi-maximum likelihood (QML), and infeasible GMM estimates of θ from the MA(1) model $y_t = e_t + \theta e_{t-1}$, where $e_t = \sigma \varepsilon_t$ and $\varepsilon_t \sim$ iid(0, 1) are generated from a generalized lambda distribution (GLD) with a skewness parameter κ_3 and no excess kurtosis. The sample size is $T = 500$, the number of Monte Carlo replications is 1000 and $\sigma = 1$. The GMM estimator is based on the moment conditions $(E(y_t y_{t-1}) - \theta\sigma^2, E(y_t^2) - (1+\theta^2)\sigma^2, E(y_t^2 y_{t-1}) - \theta^2\sigma^3\kappa_3, E(y_t^3) - (1+\theta^3)\sigma^3\kappa_3, E(y_t y_{t-1}^2) - \theta\sigma^3\kappa_3)'$. The infeasible GMM estimator is based on the same set of moment conditions but with $\sigma = 1$ assumed known. Both GMM estimators use the optimal weighting matrix based on the Newey–West HAC estimator with automatic lag selection.

that our proposed GMM estimator has properties similar to this infeasible estimator. The third is the Gaussian quasi-ML estimator of $(\theta, \sigma^2)'$ with invertibility imposed, which is used to evaluate the efficiency losses of the GMM estimator for values of θ in the invertible region ($\theta = 0.5$ and 0.7).

The results in Table 1 suggest that regardless of the degree of non-Gaussianity, the infeasible estimator produces estimates of θ that are very precise and essentially unbiased. Hence, fixing σ solves both identification problems without the need of non-Gaussianity although a prior knowledge of σ is rarely available in practice. By construction, the Gaussian QML estimator imposes invertibility and works well when the true MA parameter is in the invertible region but cannot identify the parameter values in the noninvertible region. While for $\kappa_3 = 0.35$ the identification is weak and the estimates of θ are somewhat biased, for higher values of the skewness parameter the GMM estimates of θ are practically unbiased.

Table 1 also presents the empirical probability that the particular estimator of θ is greater than or equal to one, which provides information on how often the identification of the true parameter fails. The Gaussian QML estimator is characterized by a pile-up probability at unity (which can be inferred from $P(\widehat{\theta} \geq 1)$ when $\theta_0 = 1$) as argued before. Even when $\kappa_3 = 0.35$, the GMM estimator correctly identifies if the true value of θ is in the invertible or the noninvertible region with high probability. This probability increases when $\kappa_3 = 0.85$.

Finally, to assess the accuracy of the asymptotic normality approximation in Proposition 1, Figure 1 plots the density functions of the standardized GMM estimator (t-statistic) of θ for the MA(1) model with GLD errors and a skewness parameter of 0.85 (strong identification). The sample size is $T = 3000$ and $\theta = 0.5$, 1, 1.5, and 2. Overall, the densities of the standardized GMM estimator appear to be very close to the standard normal density for all values of θ. The coverage probabilities of the 90% confidence intervals for $\theta = 0.5$, 0.7, 1, 1.5, and 2 are 91.8%, 90.5%, 92.6%, 89.5%, and 92.9%, respectively.

5. SIMULATION-BASED ESTIMATION

A caveat of the GMM estimator is that it relies on precise estimation of the higher order unconditional moments, but finite-sample biases can be nontrivial even for samples of moderate size. This can be problematic for GMM estimation of ARMA(p, q) models since a large number of higher order terms needs to be estimated. To remedy these problems, we consider the possibility of using simulation to correct for finite-sample biases (see Gourieroux, Renault, and Touzi 1999; Phillips 2012). Two estimators are considered. The first is a simulation analog of the GMM estimator, and the second is a simulated minimum distance estimator that uses auxiliary regressions to efficiently incorporate information in the higher order cumulants into a parameter vector of lower dimension. Both estimators can accommodate additional dynamics, kurtosis, and other features of the errors.

Simulation estimation of the MA(1) model was considered in Gourieroux, Monfort, and Renault (1993), Michaelides and Ng

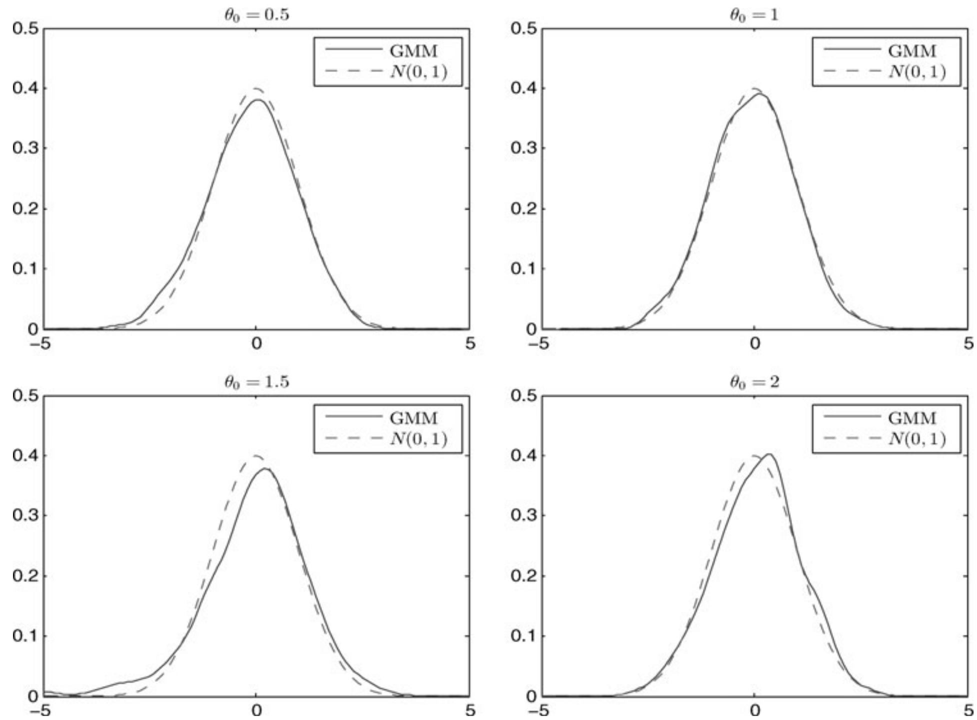

Figure 1. Density functions of the standardized GMM estimator (t-statistic) of θ based on data ($T = 3000$) generated from an MA(1) model $y_t = e_t + \theta e_{t-1}$ with $\theta = 0.5, 1, 1.5, 2$, and $e_t \sim$ iid(0, 1). The errors are drawn from a generalized lambda distribution with zero excess kurtosis and a skewness parameter equal to 0.85. For the sake of comparison, the figure also plots the standard normal ($N(0, 1)$) density.

(2000), Ghysels, Khalaf, and Vodounou (2003), and Czellar and Zivot (2008), among others, but only for the invertible case. All of these studies use an autoregression as the auxiliary model. For $\theta = 0.5$ and assuming that σ^2 is known, Gourieroux, Monfort, and Renault (1993) found that the simulation-based estimator compares favorably to the exact ML estimator in terms of bias and root mean squared error. Michaelides and Ng (2000) and Ghysels, Khalaf, and Vodounou (2003) also evaluated the properties of simulation-based estimators with σ^2 assumed known. Czellar and Zivot (2008) reported that the simulation-based estimator is relatively less biased but exhibits some instability and the tests based on it suffer from size distortions when θ_0 is close to unity (see also Tauchen 1998 for the behavior of simulation estimators near the boundary of the parameter space).

5.1 The GLD Error Simulator

The key to identification is errors with non-Gaussian features. Thus, in order for any simulation estimator to identify the parameters without imposing invertibility, we need to be able to simulate non-Gaussian errors ε_t in a flexible fashion so that y_t has the desired distributional properties.

There is evidently a large class of distributions with third and fourth moments consistent with a non-Gaussian process that one can specify. Assuming a particular parametric error distribution could compromise the robustness of the estimates. We simulate errors from the GLD $\Lambda(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ considered in Ramberg and Schmeiser (1975). This distribution has two appealing features. First, it can accommodate a wide range of values for the skewness and excess kurtosis parameters and it includes as special cases normal, log-normal, exponential, t, beta, gamma, and Weibull distributions. The second advantage is that it is easy to simulate from. The percentile function is given by

$$\Lambda(u)^{-1} = \lambda_1 + [U^{\lambda_3} + (1-U)^{\lambda_4}]/\lambda_2, \tag{15}$$

where U is a uniform random variable on [0, 1], λ_1 is a location parameter, λ_2 is a scale parameter, and λ_3 and λ_4 are shape parameters. To simulate ε_t, a U is drawn from the uniform distribution and (15) is evaluated for given values of $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. Furthermore, the shape parameters (λ_3, λ_4) and the location/scale parameters (λ_1, λ_2) can be sequentially evaluated. Since ε_t has mean zero and variance one, the parameters (λ_1, λ_2) are determined by (λ_3, λ_4) so that ε_t is effectively characterized by λ_3 and λ_4. As shown in Ramberg and Schmeiser (1975), the shape parameters (λ_3, λ_4) are explicitly related to the coefficients of skewness and kurtosis (κ_3 and κ_4) of ε_t (see the Appendix). A consequence of having to use the GLD to simulate errors is that the parameters λ_3 and λ_4 of the GLD distribution must now be estimated along with the parameters of y_t, even though these are not parameters of interest per se. In practice, these GLD parameters are identified from the higher order moments of the residuals from an auxiliary regression.

5.2 The SMM Estimator

Define the augmented parameter vector of the MA(1) model as $\gamma^+ = (\theta, \sigma^2, \lambda_3, \lambda_4)'$. Our SMM estimator is based on

$$\overline{g}(\gamma^+) = \frac{1}{T}\sum_{t=1}^{T} m_t(\gamma_0^+) - \frac{1}{TS}\sum_{t=1}^{TS} m_t^S(\gamma^+), \tag{16}$$

where $m_t(\gamma_0^+)$ is evaluated on the observed data $\mathbf{y} = (y_1, \ldots, y_T)'$ and $m_t^S(\gamma^+)$ is evaluated on the data $\mathbf{y}^S(\gamma^+) = (y_1^S, \ldots, y_T^S, \ldots, y_{TS}^S)'$ of length TS ($S \geq 1$), simulated for a candidate value of γ^+. Essentially, the quantity $m(\gamma^+)$ which is chosen to summarize the dependence of the model on the parameters γ^+ is approximated by Monte Carlo methods.

It remains to define $m_t(\gamma_0^+)$. In contrast to GMM estimation, we now need moments of the innovation errors to identify λ_3 and λ_4. The latent errors are approximated by the standardized residuals from estimation of an AR(p) model

$$y_t = \pi_0 + \pi_1 y_{t-1} + \cdots + \pi_p y_{t-p} + \sigma\epsilon_t.$$

For the MA(1) model, the moment conditions given by

$$m_t(\gamma_0^+) = \big(y_t y_{t-1} \;\; y_t^2 \;\; y_t^2 y_{t-1} \;\; y_t^3 \;\; y_t y_{t-1}^2 \;\; y_t^3 y_{t-1} \;\; y_t y_{t-1}^3 \;\; y_t^2 y_{t-1}^2 \;\; y_t^4 \;\; \hat{\epsilon}_t^3 \;\; \hat{\epsilon}_t^4\big)' \tag{17}$$

reflect information in the second-, third-, and fourth-order cumulants of the process y_t, as well as skewness and kurtosis of the errors.

To establish the consistency and asymptotic normality of the SMM estimator $\widehat{\gamma^+}$, we need some additional notation and regularity conditions. Let F_e denote the true distribution of the structural model errors and Λ_* be the class of GLDs.

Proposition 2. Consider the MA(1) model and let $G(\gamma^+) = \partial g(\gamma^+)/\partial\gamma^{+\prime}$, $\widehat{G}(\gamma^+) = \partial\overline{g}(\gamma^+)/\partial\gamma^{+\prime}$, and $\Omega = \lim_{T\to\infty} \mathrm{var}(\sqrt{T}\overline{g}(\gamma^+))$. In addition to the assumptions in Lemma 1, assume that $F_e \in \Lambda_*$, $E|e_t|^8 < \infty$, $\sup_{\gamma\in\Gamma}|\widehat{G}(\gamma^+) - G(\gamma^+)| \xrightarrow{p} 0$, γ_0^+ is in the interior of the compact parameter space Γ^+, and $\sqrt{T}\overline{g}(\gamma_0^+) \xrightarrow{d} N(0, \Omega)$. Then,

$$\sqrt{T}(\widehat{\gamma^+} - \gamma_0^+) \xrightarrow{d} N\left(0, \left(1 + \frac{1}{S}\right)\left(G(\gamma_0^+)'\Omega^{-1}G(\gamma_0^+)\right)^{-1}\right) \equiv N\big(0, \mathrm{Avar}(\widehat{\gamma^+})\big).$$

Consistency follows from identifiability of γ and the higher order cumulants play a crucial role. In our procedure, κ_3 and κ_4 are defined in terms of λ_3 and λ_4. Thus, λ_3 and λ_4 are crucial for identification of θ and σ^2 even though they are not parameters of direct interest.

A key feature of Proposition 2 is that it holds when θ is less than, equal to, or greater than one. In a Gaussian likelihood setting when invertibility is assumed for the purpose of identification, there is a boundary for the support of θ at the unit circle. Thus, the likelihood-based estimation has nonstandard properties when the true value of θ is on or near the boundary of one. In our setup, this boundary constraint is lifted because identification is achieved through higher moments instead of imposing invertibility. As a consequence, the SMM estimator $\widehat{\gamma^+}$ has classical properties provided that κ_3 and κ_4 enable identification.

Consistent estimation of the asymptotic variance of $\widehat{\gamma^+}$ can proceed by substituting a consistent estimator of Ω and evaluating the Jacobian $G(\widehat{\gamma^+})$ numerically. The computed standard errors can then be used for testing hypotheses and constructing confidence intervals. Inference on the MA parameter of interest, θ, can also be conducted by constructing confidence intervals based on inversion of the distance metric test without an explicit computation of the variance matrix $\mathrm{Avar}(\widehat{\gamma^+})$. It should be stressed that despite the choice of a flexible functional distributional form for the error simulator, our structural model is still correctly specified. This is in contrast with the semiparametric indirect inference estimator of Dridi, Guay, and Renault (2007). They considered partially misspecified structural models and thus required an adjustment in the asymptotic variance of the estimator.

5.3 The SMD Estimator

Higher order MA(q) models and general ARMA(p, q) models can in principle be estimated by GMM or SMM. But as mentioned earlier, the number of orthogonality conditions increases with p and q. Instead of selecting additional moment conditions, we combine the information in the cumulants into the auxiliary parameters that are informative about the parameters of interest. Let $\widehat{\psi}(\gamma_0^+) = \arg\min_\psi Q_T(\psi; \mathbf{y})$ and $\widetilde{\psi}^S(\gamma^+) = \arg\min_\psi Q_T(\psi; \mathbf{y}^S(\gamma^+))$ be the auxiliary parameters estimated from actual and simulated data, $Q_T(\cdot)$ denotes the objective function of the auxiliary model, and $\widehat{\Omega}$ is a consistent estimate of the asymptotic variance of $\widehat{\psi}$. Our simulated minimum distance (SMD) based on

$$\overline{g}(\gamma^+) = \widehat{\psi}(\gamma_0^+) - \widetilde{\psi}^S(\gamma^+) \tag{18}$$

shares the same asymptotic properties as the SMM estimator in Proposition 2. The SMD estimator is in the spirit of the indirect inference estimation of Gourieroux, Monfort, and Renault (1993) and Gallant and Tauchen (1996). Their estimators require that the auxiliary model is easy to estimate and that the mapping from the auxiliary parameters to the parameters of interest is well defined. We use such a mapping to collect information in the unconditional cumulants into a lower dimensional vector of auxiliary parameters to circumvent direct use of a large number of unconditional cumulants.

We consider least-square (LS) estimation of the auxiliary regressions

$$y_t = \pi_0 + \pi_1 y_{t-1} + \cdots + \pi_p y_{t-p} + \sigma\epsilon_t, \tag{19a}$$

$$y_t^2 = c_0 + c_{1,1}y_{t-1} + \cdots + c_{1,r}y_{t-r} + c_{2,1}y_{t-1}^2 + \cdots + c_{2,r}y_{t-r}^2 + v_t \tag{19b}$$

with an appropriate choice of p and r. Equation (19a) has been used in the literature for simulation estimation of MA(1) models when invertibility is imposed, and often with σ^2 assumed known. We complement (19a) with the regression defined in (19b). The parameters of this regression parsimoniously summarize information in the higher moments of the data. Compared to the SMM in which the auxiliary parameters are unconditional moments, the auxiliary parameters ψ are based on conditional moments. Equation (19b) also provides a simple check for the prerequisite for identification. If the c coefficients are jointly zero, identification would be in jeopardy.

Let $\widehat{\kappa}_3$ and $\widehat{\kappa}_4$ denote the sample third and fourth moments of the ordinary LS (OLS) residuals in (19a). The auxiliary

Table 2. SMM and SMD estimates of θ from MA(1) model with asymmetric errors

	SMM				SMD			
	Mean	Med.	$P(\widehat{\theta} \geq 1)$	Std.	Mean	Med.	$P(\widehat{\theta} \geq 1)$	Std.
GLD, $\sigma_t = \sigma$								
$\theta_0 = 0.5$	0.488	0.484	0.001	0.054	0.503	0.503	0.000	0.043
$\theta_0 = 0.7$	0.693	0.688	0.000	0.083	0.705	0.703	0.002	0.053
$\theta_0 = 1.0$	0.949	0.988	0.421	0.137	0.973	0.982	0.406	0.089
$\theta_0 = 1.5$	1.563	1.520	0.962	0.280	1.482	1.493	0.980	0.104
$\theta_0 = 2.0$	1.903	1.959	0.940	0.337	1.996	1.995	0.988	0.180
GLD + ARCH								
$\theta_0 = 0.5$	0.437	0.426	0.013	0.064	0.549	0.479	0.062	0.055
$\theta_0 = 0.7$	0.648	0.636	0.006	0.099	0.748	0.687	0.143	0.059
$\theta_0 = 1.0$	0.929	0.959	0.318	0.160	1.068	1.087	0.790	0.117
$\theta_0 = 1.5$	1.573	1.561	0.940	0.274	1.483	1.486	0.983	0.113
$\theta_0 = 2.0$	1.861	1.956	0.883	0.374	1.926	1.924	0.978	0.240

NOTES: The table reports the mean, median (med.), probability that $\widehat{\theta} \geq 1$, and standard deviation (std.) of the SMM estimates of θ from the MA(1) model $y_t = e_t + \theta e_{t-1}$, where $e_t = \sigma_t \varepsilon_t$, $\varepsilon_t \sim$ iid(0, 1) are generated from a generalized lambda distribution (GLD) with a skewness parameter $\kappa_3 = 0.85$ (and no excess kurtosis) and $\sigma_t = \sigma = 1$ or $\sigma_t^2 = 0.7 + 0.3e_{t-1}^2$. The sample size is $T = 500$ and the number of Monte Carlo replications is 1000. The SMM estimator is based on the moment conditions $m_{\mathrm{SMM},t}$, defined in (17), and the SMD estimator is based on the auxiliary parameter vector ψ_{SMD}, defined in (19c). The SMM and SMD estimators use the optimal weighting matrix based on the Newey–West HAC estimator.

parameter vector based on the data is

$$\widehat{\psi}(\gamma_0^+) = \left(\widehat{\pi}_0, \widehat{\pi}_1, \ldots, \widehat{\pi}_p, \widehat{c}_0, \widehat{c}_{1,1}, \ldots, \widehat{c}_{1,r}, \widehat{c}_{2,1}, \ldots, \widehat{c}_{2,r}, \widehat{\kappa}_3, \widehat{\kappa}_4\right)'. \tag{19c}$$

The parameter vector $\psi^S(\gamma^+)$ is analogously defined, except that the auxiliary regressions are estimated with data simulated for a candidate value of γ.

5.4 Finite-Sample Properties of the Simulation-Based Estimators

To implement the SMM and SMD estimators, we simulate TS errors from the generalized lambda error distribution. Larger values of S (the number of simulated sample paths of length T) tend to smooth the objective functions, which improves the identification of the MA parameter. As a result, we set $S = 20$

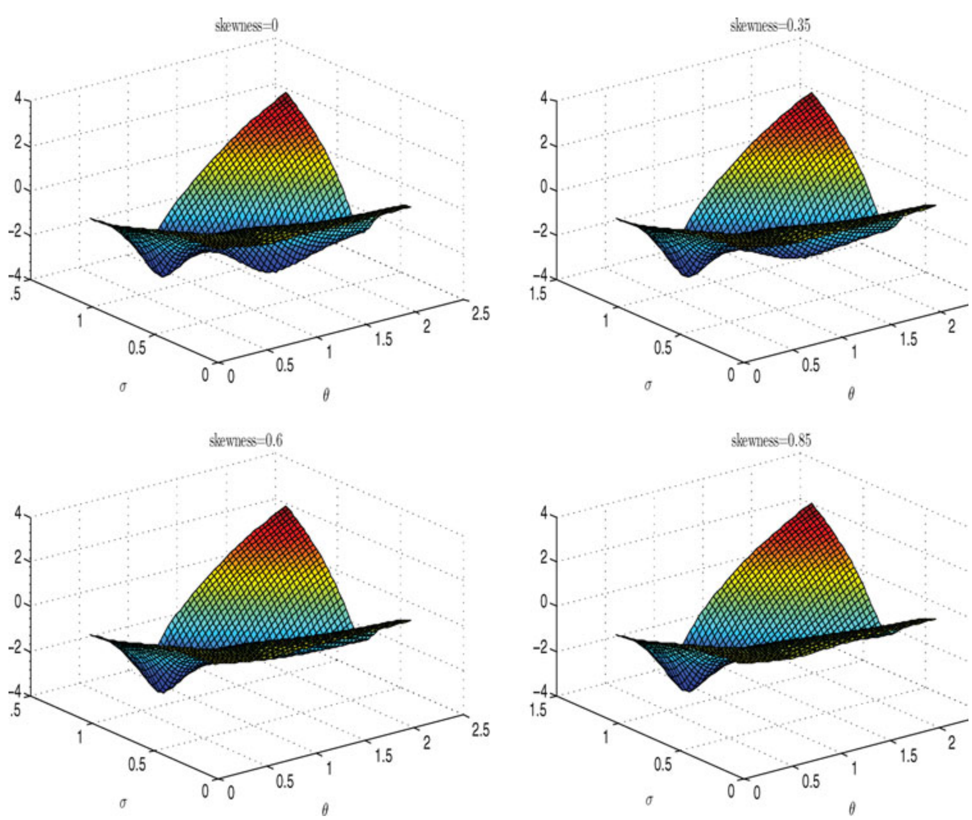

Figure 2. Logarithm of the objective function of simulation-based estimator of θ and σ based on data ($T = 1000$) generated from an MA(1) model $y_t = e_t + \theta e_{t-1}$ with $\theta = 0.7$ and $e_t \sim$ iid(0, 1). The errors are drawn from a generalized lambda distribution with zero excess kurtosis and a skewness parameter equal to 0, 0.35, 0.6, and 0.85.

Table 3. SMD, SMM, and Gaussian QML estimates of θ and α from an ARMA(1, 1) model with exponential/mixture of normals errors

	θ				α		
Errors/Estimator	Mean	Med.	Std.	$P(\lvert\widehat{\theta}\rvert \geq 1)$	Mean	Med.	Std.
Exponential errors							
		$\theta_0 = -1.5$				$\alpha_0 = 0.5$	
SMD	−1.552	−1.489	0.544	0.954	0.493	0.501	0.162
SMM	−1.497	−1.480	0.378	0.994	0.496	0.504	0.109
Gaussian QML	−0.652	−0.686	0.206	0.000	0.482	0.511	0.217
		$\theta_0 = -2$				$\alpha_0 = 0.5$	
SMD	−2.039	−2.001	0.626	0.976	0.483	0.490	0.134
SMM	−1.919	−1.958	0.648	0.967	0.473	0.501	0.194
Gaussian QML	−0.011	0.010	0.571	0.000	0.011	−0.003	0.567
Mixture errors							
		$\theta_0 = -1.5$				$\alpha_0 = 0.5$	
SMD	−1.501	−1.480	0.415	0.967	0.505	0.516	0.137
SMM	−1.277	−1.379	0.671	0.805	0.444	0.512	0.319
Gaussian QML	−0.660	−0.688	0.168	0.000	0.487	0.510	0.186
		$\theta_0 = -2$				$\alpha_0 = 0.5$	
SMD	−1.728	−1.723	0.498	0.978	0.570	0.580	0.191
SMM	−1.537	−1.678	1.015	0.785	0.457	0.516	0.380
Gaussian QML	−0.012	−0.003	0.563	0.000	0.009	0.001	0.558

NOTES: The table reports the mean, median (med.), standard deviation (std.), and the probability that $P(|\widehat{\theta}| \geq 1)$ of the SMD, SMM, and Gaussian QML estimates of θ and α from the ARMA(1, 1) model $(1 - \alpha L)y_t = (1 + \theta L)e_t$, where $e_t = \sigma\varepsilon_t$ and ε_t is an exponential random variable with a scale parameter equal to one (exponential errors) or a mixture of normals random variable with mixture probabilities 0.1 and 0.9, means −0.9 and 0.1, and standard deviations 2 and 0.752773, respectively (mixture errors). The exponential errors are recentered and rescaled to have mean zero and variance one. The sample size is $T = 500$ and the number of Monte Carlo replications is 1000.

although $S > 20$ seems to offer even further improvement, especially for small T, but at the cost of increased computational time. The SMM and SMD estimators both use $p = 4$. SMD additionally assumes $r = 1$ in the auxiliary model (19b).

As is true of all nonlinear estimation problems, the numerical optimization problem must take into account the possibility of local minima, which arises when the invertibility condition is not imposed. Thus, the estimation always considers two sets of initial values. Specifically, we draw two starting values for θ—one from a uniform distribution on (0, 1) and one from a uniform distribution on (1, 2)—with the starting value for σ set equal to $\sqrt{\widehat{\sigma}_y^2/(1+\theta^2)}$ for each of the starting values for θ. The starting values for the shape parameters of the GLD λ_3 and λ_4 are set equal to those of the standard normal distribution (with $\kappa_3 = 0$ and $\kappa_4 = 3$). In this respect, the starting values of θ, σ, λ_3, and λ_4 contain little prior knowledge of the true parameters.

MA(1). In the first experiment, data are generated from

$$y_t = e_t + \theta e_{t-1}, \quad e_t = \sigma_t \varepsilon_t,$$

where $\varepsilon_t \sim \text{iid}(0, 1)$ is drawn from a GLD with zero excess kurtosis and a skewness parameter 0.85 with (i) $\sigma_t = \sigma = 1$ or (ii) $\sigma_t = 0.7 + 0.3e_{t-1}^2$ (ARCH errors). The sample size is $T = 500$, the number of Monte Carlo replications is 1000 and θ takes the values of 0.5, 0.7, 1, 1.5, and 2. Note that the structural model used for SMM and SMD does not impose the ARCH structure of the errors, that is, the error distribution is misspecified. This case is useful for evaluating the robustness properties of the proposed SMM and SMD estimators.

Table 2 reports the mean and median estimates of θ, the standard deviation of the estimates for which identification is achieved and the probability that the estimator is equal to or greater than one. When the errors are iid drawn from the GLD distribution, the SMM estimator of θ exhibits only a small bias for some values of θ (e.g., $\theta_0 = 2$). While there is a positive probability that the SMM estimator will converge to $1/\theta$ instead of θ (especially when θ is in the noninvertible region), this probability is fairly small and it disappears completely for larger T (not reported to conserve space). When the error distribution is misspecified (GLD errors with ARCH structure), the properties of the estimator deteriorate (the estimator exhibits a larger bias) but the invertible/noninvertible values of θ are still identified with high probability. However, the SMD estimator provides a substantial bias correction, efficiency gain, and identification improvement. Interestingly, in terms of precision, the SMD estimator appears to be more efficient than the infeasible estimator in Table 1 for values of θ in the invertible region. The SMD estimator continues to perform well even when the error simulator is misspecified.

Figure 2 illustrates how identification depends on skewness by plotting the log of the objective function for the SMD estimator averaged over 1000 Monte Carlo replications of the MA(1) model with $\theta = 0.7$ and $\sigma = 1$. The errors are generated from GLD with zero excess kurtosis and three values of the skewness parameter: 0, 0.35, 0.6, and 0.85. In evaluating the objective function, the values of the lambda parameters in the GLD are set equal to their true values. The first case (no skewness) corresponds to lack of identification and there are two pronounced local minima at θ and $1/\theta$. As the skewness of the error distribution increases, the second local optima at $1/\theta$ flattens out and it almost completely disappears when the error distribution is highly asymmetric.

ARMA(1, 1). In the second simulation experiment, data are generated according to

$$y_t = \alpha y_{t-1} + e_t + \theta e_{t-1}, \quad (20)$$

where e_t is (i) a standard exponential random variable with a scale parameter equal to one, which is recentered and rescaled to have mean zero and variance 1 or (ii) a mixture of normals random variable with mixture probabilities 0.1 and 0.9, means −0.9 and 0.1, and standard deviations 2 and 0.752773, respectively. The second error distribution is included to assess the robustness properties of the simulation-based estimator to error distributions that are not members of the GLD family.

We consider two parameterizations that give rise to a causal process with a noninvertible MA component. The first parameterization is $\alpha = 0.5$ and $\theta = -1.5$. The second parameterization, $\alpha = 0.5$ and $\theta = -2$, produces an all-pass ARMA(1, 1) process, which is characterized by $\theta = -1/\alpha$. This all-pass process possesses some interesting properties (see Davis 2010). First, y_t is uncorrelated but is conditionally heteroscedastic. Second, if one imposes invertibility by letting $\theta = -\alpha$ and scale up the error variance by $(1/\alpha)^2$, the process is iid and the AR and MA parameters are not separately identifiable. Imposing invertibility in such a case is not innocuous, and estimation of the parameters of this model is quite a challenging task.

Table 3 presents the finite-sample properties of the SMD and SMM estimators for the ARMA(1, 1) model in (20) using the same auxiliary parameters and moment conditions for the estimation of MA(1). For comparison, we also include the Gaussian quasi-ML estimator. The SMD estimates of θ appear unbiased for the exponential distribution and are somewhat downward biased for the mixture of normals errors. But, overall, the SMD estimator identifies correctly the AR and MA components with high probability. The performance of the SMM estimator is also satisfactory but it is dominated by the SMD estimator. The Gaussian QML estimator imposes invertibility and completely fails to identify the AR and MA parameters when $\alpha = 0.5$ and $\theta = -2$. Even with a misspecified error distribution and a fairly parsimonious auxiliary model, the finite-sample properties of our proposed simulation-based estimators remain quite attractive.

5.5 Empirical Application: 25 Fama-French Portfolio Returns

Noninvertibility can be consistent with economic theory. For example, suppose $y_t = E_t \sum_{s=0}^{\infty} \delta^s x_{t+s}$ is the present value of $x_t = e_t + \omega e_{t-1}$. As shown by Hansen and Sargent (1991), the solution $y_t = (1 + \delta\omega)e_t + \omega e_{t-1} = h(L)e_t$ implies that the root of $h(z)$ is $-\frac{1+\delta\omega}{\omega}$, which can be on or inside the unit circle even if $|\omega| < 1$. If there is no discounting and $\delta = 1$, y_t has an MA unit root when $\omega = -0.5$ and $h(L)$ is noninvertible in the past whenever $\omega < -0.5$. Note that even if an autoregressive processes is causal, it is still possible for the roots of $h(L) = \frac{\delta\omega(\delta) - L\omega(L)}{\delta - L}$ to be inside the unit disk.

Present value models are used to analyze variables with a forward looking component including stock and commodity prices. We estimate an MA(1) model for each of the 25 Fama-French portfolio returns using the Gaussian QML and the proposed SMM and SMD estimators. The data are monthly returns on the value-weighted 25 Fama-French size and book-to-market ranked portfolios from January 1952 until August 2013 (from Kenneth French's website). The portfolios are the intersections of five portfolios formed on size (market equity) and five portfolios formed on the ratio of book equity to market equity. The size (book-to-market) breakpoints are the NYSE quintiles and are denoted by "small, 2, 3, 4, big" ("low, 2, 3, 4, high") in Table 4.

Table 4 presents the sample skewness and kurtosis as well as the estimates and the corresponding standard errors (in parentheses below the estimate) for each estimator and portfolio return. All of the returns exhibit some form of non-Gaussianity, which is necessary for identifying possible noninvertible MA components. The Gaussian QML produces estimates of the MA coefficient that are small but statistically significant (with a few exceptions in the "big" size category). The SMM relaxes the invertibility constraint and delivers somewhat higher estimates of the MA parameter but most of these estimates still fall in the invertible region. By contrast, the SMD estimator suggests that all of the 25 Fama-French portfolio returns appear to be driven

Table 4. SMD, SMM, and Gaussian QML estimates of MA(1) model for stock portfolio returns

		Skewness	Kurtosis	QML	SMM	SMD
	Low	0.039	5.244	0.155 (0.028)	4.711 (0.650)	4.325 (0.470)
	2	0.030	6.136	0.160 (0.027)	0.273 (0.028)	4.043 (0.417)
Small	3	−0.132	5.889	0.179 (0.032)	0.287 (0.027)	3.802 (0.348)
	4	−0.164	6.131	0.180 (0.034)	4.754 (0.538)	4.092 (0.455)
	High	−0.208	6.464	0.241 (0.032)	3.368 (0.349)	2.944 (0.254)
	Low	−0.318	4.677	0.144 (0.032)	0.212 (0.027)	3.694 (0.289)
	2	−0.419	5.551	0.143 (0.035)	0.219 (0.697)	3.880 (0.384)
2	3	−0.458	6.105	0.153 (0.035)	0.251 (0.026)	3.763 (0.312)
	4	−0.439	6.148	0.156 (0.035)	0.241 (0.025)	4.120 (0.394)
	High	−0.414	6.186	0.166 (0.030)	0.232 (0.027)	3.745 (0.306)
	Low	−0.371	4.701	0.117 (0.030)	0.178 (0.022)	3.001 (0.162)
	2	−0.506	5.936	0.151 (0.035)	0.278 (0.022)	3.702 (0.323)
3	3	−0.510	5.324	0.146 (0.034)	4.884 (0.386)	3.553 (0.272)
	4	−0.276	5.314	0.142 (0.034)	0.246 (0.026)	3.537 (0.283)
	High	−0.305	6.081	0.154 (0.033)	4.981 (0.488)	3.875 (0.340)
	Low	−0.234	4.933	0.104 (0.033)	0.168 (0.022)	3.338 (0.180)
	2	−0.585	6.135	0.143 (0.034)	0.203 (0.022)	3.649 (0.416)
4	3	−0.503	6.348	0.140 (0.032)	0.264 (0.023)	3.682 (0.354)
	4	−0.231	4.930	0.092 (0.035)	0.212 (0.020)	4.045 (0.300)
	High	−0.193	5.385	0.118 (0.032)	0.244 (0.021)	4.680 (0.405)
	Low	−0.253	4.565	0.065 (0.030)	0.106 (0.028)	5.078 (0.799)
	2	−0.362	4.677	0.052 (0.034)	0.157 (0.024)	5.602 (0.686)
Big	3	−0.264	5.209	0.035 (0.031)	0.100 (0.029)	6.457 (1.119)
	4	−0.168	4.608	0.025 (0.032)	0.125 (0.021)	6.206 (1.140)
	High	−0.200	4.002	0.072 (0.032)	0.140 (0.020)	4.803 (0.611)

NOTES: The table reports the SMD, SMM, and Gaussian quasi-ML estimates and standard errors (in parentheses below the estimates) for the MA(1) model $y_t = e_t + \theta e_{t-1}$, where $e_t \sim \text{iid}(0, \sigma^2)$ and y_t is one of the 25 Fama-French portfolio returns. The first two columns report the sample skewness and kurtosis of y_t. The standard errors for SMM and SMD are constructed using the asymptotic approximation in Proposition 2.

by a noninvertible MA component. The results are consistent with the finding through simulations that the SMD is more capable of estimating θ in the correct invertibility space. The SMD estimates are fairly stable across the different portfolio returns with a slight increase in their magnitude and standard errors for the "big" size portfolios. Also, a higher precision of the MA estimates is typically associated with returns that are characterized by larger departures from Gaussianity. Overall, our SMD method provides evidence in support of noninvertibility in stock returns.

6. CONCLUSIONS

This article proposes classical and simulation-based GMM estimation of possibly noninvertible MA models with non-Gaussian errors. The identification of the structural parameters is achieved by exploiting the non-Gaussianity of the process through third-order cumulants. This type of identification also removes the boundary problem at the unit circle, which gives rise to the pile-up probability and nonstandard asymptotics of the Gaussian maximum likelihood estimator. As a consequence, the proposed GMM estimators are root-T consistent and asymptotically normal over the whole parameter range, provided that the non-Gaussianity in the data is sufficiently large to ensure identification.

Other research questions arise once the assumption of invertibility is relaxed. A potential problem with the GMM estimator is that the number of orthogonality conditions can be quite large. This is especially problematic for ARMA(p, q) models. Ideally, the orthogonality conditions should be selected or weighted in an optimal fashion. More generally, how to determine the lag length of the heteroskedasticity and autocorrelation consistent (HAC) estimator without imposing invertibility remains a topic for future research.

APPENDIX: PROOFS

Proof of Lemma 1. The result in part (a) follows immediately by noticing that $g(\gamma_1)$ and $g(\gamma_2)$, where $g = (E(y_t y_{t-1}), E(y_t^2), E(y_t^2 y_{t-1}))'$, are observationally equivalent for $\gamma_1 = (\theta, \sigma^2, \kappa_3)'$ and $\gamma_2 = (1/\theta, \theta^2\sigma^2, \theta\kappa_3)'$. For part (b), let us define the derivative matrix of $g(\gamma) = (g_2', g_3')'$ as

$$G = \begin{pmatrix} \sigma^2 & \theta & 0 \\ 2\theta\sigma^2 & (1+\theta^2) & 0 \\ 2\theta\sigma^3\kappa_3 & \frac{3}{2}\theta^2\sigma\kappa_3 & \theta^2\sigma^3 \\ 3\theta^2\sigma^3\kappa_3 & \frac{3}{2}(1+\theta^3)\sigma\kappa_3 & (1+\theta^3)\sigma^3 \\ \sigma^3\kappa_3 & \frac{3}{2}\theta\sigma\kappa_3 & \theta\sigma^3 \end{pmatrix}$$

with $G_{[1,2,i]}$ for $i = 3, 4$, or 5 denoting its corresponding 3×3 block. Direct calculations of the determinants give $|G|_{[1,2,3]} = (1-\theta^2)\theta^2\sigma^5$, $|G|_{[1,2,4]} = (1-\theta^2)(1+\theta^3)\sigma^5$, and $|G|_{[1,2,5]} = (1-\theta^2)\theta\sigma^5$, which are all zero at $|\theta| = 1$.

Proof of Lemma 2. Giannakis and Mendel (1989) solved $\beta(\gamma)$ from the system of overdetermined equations but did not establish uniqueness of the solution. The argument for identification from third- and fourth-order cumulants (i.e., Equation (7)) and fourth-order cumulants (i.e., Equation (9)) are similar. We begin with (7).

Identification of MA(q) Models Using Third-Order Cumulants. The system of equations can be expressed as $A\beta(\gamma) = b$, where

$$A = \begin{bmatrix} B & C \\ D & E \end{bmatrix}, \quad C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix},$$

where

$$B = \begin{bmatrix} 0 & 0 & \dots & 0 \\ c_3(-q) & 0 & \dots & 0 \\ c_3(-q+1) & c_3(-q) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ c_3(q-1) & c_3(q-2) & \dots & c_3(0) \end{bmatrix},$$

$$E = \begin{bmatrix} 0 & -c_2(q) & -c_2(q-1) & \cdots & -c_2(1) \\ 0 & 0 & -c_2(q) & \cdots & -c_2(2) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \cdots & -c_2(q) \end{bmatrix},$$

$$D = \begin{bmatrix} c_3(q) & c_3(q-1) & \cdots & c_3(1) \\ 0 & c_3(q) & \cdots & c_3(2) \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \cdots & c_3(q) \end{bmatrix},$$

$$C_1 = \begin{bmatrix} -c_2(q) & 0 & 0 & \cdots & 0 \\ -c_2(q-1) & -c_2(q) & 0 & \cdots & 0 \\ -c_2(q-2) & -c_2(q-1) & -c_2(q) & \cdots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ -c_2(0) & -c_2(1) & -c_2(2) & \cdots & -c_2(q) \end{bmatrix},$$

$$C_2 = \begin{bmatrix} -c_2(1) & -c_2(0) & -c_2(1) & \cdots & -c_2(q-1) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ -c_2(q) & -c_2(q-1) & -c_2(q-2) & \cdots & -c_2(0) \end{bmatrix}.$$

The rank of the $(3q+1) \times (2q+1)$ matrix A is the sum of the column rank of the submatrix consisting of B and D, and the rank of the submatrix consisting of C and E. The rank of the first subblock is determined by the rank of the $q \times q$ square matrix D, which is q if $c_3(q) \neq 0$. The rank of C is determined by the rank of the square matrix C_1, which is $(q+1)$ if $c_2(q) \neq 0$. The full rank result follows from the assumption that $c_2(q)$ and $c_3(q)$ are nonzero.

Since $c_2(q) \neq 0$ and $c_3(q) \neq 0$ from the assumptions of Lemma 2, the triangular matrices C_1 and D have column ranks of $q+1$ and q, respectively. Therefore, A has a full column rank of $2q+1$ and the parameter vector $\beta(\gamma)$ can be obtained as a unique solution to the system of Equation (8). Since the derivative matrix of $\beta(\gamma)$ given by

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 1 \\ 2\kappa_3\sigma\theta_1 & 0 & \cdots & 0 & \theta_1^2 \\ 0 & 2\kappa_3\sigma\theta_2 & \cdots & 0 & \theta_2^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \kappa_3\sigma\theta_q & \theta_q^2 \end{bmatrix}$$

is of full column rank, the parameter vector of interest $\gamma = (\theta_1, \dots, \theta_q, \kappa_3\sigma)'$ is identifiable.

Identification of MA(q) Models Using Fourth-Order Cumulants. The MA(q) model implies the following relation between the diagonal slices of the fourth-order cumulants and the $q+1$ vector of parameters $\gamma = (\theta_1, \ldots, \theta_q, \sigma^2(\kappa_4 - 3))'$:

$$\sum_{i=1}^{q} \theta_i c_4(\tau - i) - \sigma^2(\kappa_4 - 3) \sum_{i=0}^{q} \theta_i^3 c_2(\tau - i) + c_4(\tau) = 0, \quad -q \le \tau \le 2q. \tag{A.1}$$

Define

$$\beta(\gamma) = \big(\theta_1, \ldots, \theta_q, \sigma^2(\kappa_4 - 3), \sigma^2(\kappa_4 - 3)\theta_1^3, \ldots, \sigma^2(\kappa_4 - 3)\theta_q^3\big)',$$
$$b = [-c_4(-q) \;\; -c_4(-q+1) \;\; \cdots \;\; -c_4(0) \;\; -c_4(1) \;\; \cdots \;\; -c_4(q-1) \;\; -c_4(q) \;\; 0 \;\; 0 \;\; \cdots \;\; 0]',$$

and

$$A = \begin{bmatrix} 0 & 0 & \ldots & 0 & -c_2(q) & 0 & 0 & \cdots & 0 \\ c_4(-q) & 0 & \ldots & 0 & -c_2(q-1) & -c_2(q) & 0 & \cdots & 0 \\ c_4(-q+1) & c_4(-q) & \ldots & 0 & -c_2(q-2) & -c_2(q-1) & -c_2(q) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ c_4(q-1) & c_4(q-2) & \ldots & c_4(0) & -c_2(q) & -c_2(q-1) & -c_2(q-2) & \cdots & -c_2(0) \\ c_4(q) & c_4(q-1) & \cdots & c_4(1) & 0 & -c_2(q) & -c_2(q-1) & \cdots & -c_2(1) \\ 0 & c_4(q) & \cdots & c_4(2) & 0 & 0 & -c_2(q) & \cdots & -c_2(2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & c_4(q) & 0 & 0 & 0 & \cdots & -c_2(q) \end{bmatrix}.$$

Then, the system of Equation (A.1) can be expressed as

$$A\beta(\gamma) = b$$

and the identification of $\beta(\gamma)$ and γ follows similar arguments as those for the third-order cumulants.

Proof of Lemma 3. The proof follows some of the arguments in the proof of Theorem 1 in Tugnait (1995). Consider two ARMA (p, q) models $\alpha_1(L)y_t = \theta_1(L)e_t$ and $\alpha_2(L)y_t = \theta_2(L)e_t$, which can be rewritten as $z_t = \alpha_2(L)\theta_1(L)e_t = \Theta_1(L)e_t$ and $z_t = \alpha_1(L)\theta_2(L)e_t = \Theta_2(L)e_t$, where $z_t = a_1(L)a_2(L)y_t$. Let $\phi_1 = (\alpha_{1,1}, \ldots, \alpha_{1,p}, \theta_{1,1}, \ldots, \theta_{1,q})'$ and $\phi_2 = (\alpha_{2,1}, \ldots, \alpha_{2,p}, \theta_{2,1}, \ldots, \theta_{2,q})'$. Note that z_t is an MA$(p+q)$ process since $\Theta_1(\cdot) = \alpha_2(L)\theta_1(L)$ and $\Theta_2(\cdot) = \alpha_1(L)\theta_2(L)$ are polynomials of order $p+q$. As in Lemma 2, we can write

$$A\beta_1(\phi_1, \kappa_3\sigma) = b$$
$$A\beta_2(\phi_2, \kappa_3\sigma) = b,$$

where A and b are functions of second and third cumulants of z_t. But from Lemma 2, there exists a unique solution to the system of equations $A\beta(\phi, \kappa_3\sigma) = b$. Hence, there is a one-to-one mapping between (A, b) and $\beta(\phi, \kappa_3\sigma)$ and the two ARMA models are identical in the sense that $\phi_1 = \phi_2$. Therefore, $\phi = (\alpha_1, \ldots, \alpha_p, \theta_1, \ldots, \theta_q)'$ is identifiable from the second and third cumulants used in constructing A and b, provided that $c_2(p+q) \ne 0$ and $c_3(p+q) \ne 0$.

Proof of Proposition 1. The results in Section 3 ensure global and local identifiability of γ_0. The consistency of $\widehat{\gamma}$ follows from the identifiability of γ_0 and the compactness of Γ. Taking a mean value expansion of the first-order conditions of the GMM problem and invoking the central limit theorem deliver the desired asymptotic normality result.

The GLD Distribution. The two parameters λ_3, λ_4 are related to κ_3 and κ_4 as follows (see Ramberg and Schmeiser 1975):

$$\kappa_3 = \frac{c - 3ab + 2a^3}{\lambda_2^3},$$
$$\kappa_4 = \frac{d - 4ac + 6a^2b - 3a^4}{\lambda_2^4},$$

where $a = \frac{1}{1+\lambda_3} - \frac{1}{1+\lambda_4}$, $b = \frac{1}{1+2\lambda_3} + \frac{1}{1+2\lambda_4} - 2\text{Beta}(1+\lambda_3, 1+\lambda_4)$, $\lambda_2 = \sqrt{B - A^2}$, $c = \frac{1}{1+3\lambda_3} - 3\text{Beta}(1+2\lambda_3, 1+\lambda_4) + 3\text{Beta}(1+\lambda_3, 1+2\lambda_4) - \frac{1}{1+3\lambda_4}$, $d = \frac{1}{1+4\lambda_3} - 4\text{Beta}(1+3\lambda_3, 1+\lambda_4) + 6\text{Beta}(1+2\lambda_3, 1+2\lambda_4) - 4\text{Beta}(1+\lambda_3, 1+3\lambda_4) + \frac{1}{1+4\lambda_4}$, and $\text{Beta}(\cdot, \cdot)$ denotes the beta function.

ACKNOWLEDGMENTS

The authors thank the Editor, an Associate Editor, two anonymous referees, Prosper Dovonon, Anders Bredahl Kock, Ivana Komunjer, and the participants at the CESG meeting at Queen's University for useful comments and suggestions. The second author acknowledges financial support from the National Science Foundation (SES-0962431). The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

[Received March 2013. Revised May 2014.]

References

Akaike, H. (1966), "Note on Higher Order Spectra," *Annals of the Institute of Statistical Mathematics*, 18, 123–126. [407]

Anderson, T. W., and Takemura, A. (1986), "Why Do Noninvertible Estimated Moving Average Models Occur?" *Journal of Time Series Analysis*, 7, 235–254. [405]

Andrews, B., Davis, R., and Breidt, F. J. (2006), "Maximum Likelihood Estimation of All-Pass Time Series Models," *Journal of Multivariate Analysis*, 97, 1638–1659. [403]

——— (2007), "Rank-Based Estimation of All-Pass Time Series Models," *The Annals of Statistics*, 35, 844–869. [403]

Brockwell, P. J., and Davies, R. A. (1991), *Time Series Theory and Methods* (2nd ed.), New York: Springer-Verlag. [404]

Czellar, V., and Zivot, E. (2008), "Improved Small Sample Inference for Efficient Method of Moments and Indirect Inference Estimators," Mimeo, University of Washington. [410]

Davis, R. (2010), "All-Pass Processes With Applications to Finance," 7th International Iranian Workshop on Stochastic Processes, Tehran, Iran (Plenary Talk), Nov. 30–Dec. 2, 2010. [414]

Davis, R., and Dunsmuir, W. (1996), "Maximum Likelihood Estimation for MA(1) Processes With a Root on the Unit Circle," *Econometric Theory*, 12, 1–20. [405,406]

Davis, R., and Song, L. (2011), "Unit Roots in Moving Averages Beyond First Order," *The Annals of Statistics*, 39, 3062–3091. [405]

Dovonon, P., and Renault, E. (2013), "Testing for Common Conditionally Heteroskedastic Factors," *Econometrica*, 81, 2561–2586. [405]

Dridi, R., Guay, A., and Renault, E. (2007), "Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models," *Journal of Econometrics*, 136, 397–430. [411]

Duffie, D., and Singleton, K. (1993), "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–952. [404]

Dufour, J.-M., and Hsiao, C. (2008), "Identification," in *New Palgrave Dictionary of Economics* (2nd ed.), eds. S. N. Durlauf and L. E. Blume, London: Palgrave Macmillan Press. [405]

Fernández-Villaverde, F., Rubio-Ramírez, J., Sargent, T., and Watson, M. (2007), "A, B, Cs and (D)'s for Understanding VARs," *American Economic Review*, 97, 1021–1026. [403]

Fisher, F. M. (1961), "Identifiability Criteria in Non-Linear Systems," *Econometrica*, 29, 574–590. [405]

——— (1965), "Identifiability Criteria in Non-Linear Systems: A Further Note," *Econometrica*, 33, 197–206. [405]

Friedlander, B., and Porat, B. (1990), "Asymptotically Optimal Estimation of MA and ARMA Parameters of Non-Gaussian Processes of Higher-Order Moments," *IEEE Transactions on Automatic Control*, 35, 27–35. [407,408]

Gallant, R., and Tauchen, G. (1996), "Which Moments to Match," *Econometric Theory*, 12, 657–681. [404,411]

Ghysels, E., Khalaf, L., and Vodounou, C. (2003), "Simulation Based Inference in Moving Average Models," *Annales D' Economie et Statistique*, 69, 85–99. [410]

Giannakis, G., and Mendel, J. (1989), "Identification of Nonminimum Phase Systems Using Higher Order Statistics," *IEEE Transactions, Acoustics, Speech, and Signal Processing*, 37, 360–377. [406,407,415]

Giannakis, G., and Swami, A. (1992), "Identifiability of General ARMA Processes Using Linear Cumulant-Based Estimators," *Automatica*, 28, 771–779. [406,407]

Gospodinov, N. (2002), "Bootstrap Based Inference in Models With a Nearly Noninvertible Moving Average Component," *Journal of Business and Economic Statistics*, 20, 254–268. [405]

Gourieroux, C., Monfort, A., and Renault, E. (1993), "Indirect Inference," *Journal of Applied Econometrics*, 85, 85–118. [404,409,411]

Gourieroux, C., Renault, E., and Touzi, N. (1999), "Calibration by Simulation for Small Sample Bias Correction," *in Simulation-Based Inference in Econometrics: Methods and Applications*, eds. R. Mariano, T. Schuermann, and M. Weeks, Cambridge: Cambridge University Press, pp. 328–358. [409]

Hansen, L., and Sargent, T. (1991), "Two Difficulties in Interpreting Vector Autoregressions," in *Rational Expectations Econometrics*, eds. L. P. Hansen, and T. J. Sargent, London: Westview, pp. 77–119. [403,414]

Huang, J., and Pawitan, Y. (2000), "Quasi-Likelihood Estimation of Non-Invertible Moving Average Processes," *Scandinavian Journal of Statistics*, 27, 689–702. [403]

Komunjer, I. (2012), "Global Identification in Nonlinear Models With Moment Restrictions," *Econometric Theory*, 28, 719–729. [405]

Komunjer, I., and Ng, S. (2011), "Dynamic Identification of Dynamic Stochastic General Equilibrium Models," *Econometrica*, 79, 1995–2032. [403]

Lii, K., and Rosenblatt, M. (1982), "Deconvolution and Estimation of Transfer Function Phase Coefficients for Non-Gaussian Linear Processes," *The Annals of Statistics*, 10, 1195–1208. [405,406]

——— (1992), "An Approximate Maximum Likelihood Estimation of Non-Gaussian Non-Minimum Phase Moving Average Processes," *Journal of Multivariate Analysis*, 43, 272–299. [403]

Lippi, M., and Reichlin, L. (1993), "The Dynamic Effects of Aggregate Demand and Supply Disturbances: Comment," *American Economic Review* 83, 644–652. [403]

Meitz, M., and Saikkonen, P. (2011), "Maximum Likelihood Estimation of a Non-Invertible ARMA Model With Autoregressive Conditional Heteroskedasticity," Mimeo, University of Helsinki. [404]

Mendel, J. M. (1991), "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications," *Proceedings of the IEEE*, 79, 278–305. [406,407]

Michaelides, A., and Ng, S. (2000), "Estimating the Rational Expectations Model of Speculative Storage: A Monte Carlo Comparison of Three Simulation Estimators," *Journal of Econometrics*, 96, 231–266. [410]

Na, Y., Kim, K., Song, I., and Kim, T. (1995), "Identification of Nonminimum Phase FIR Systems Using the Third and Fourth Order Cumulants," *IEEE Transactions on Signal Processing*, 43, 2018–2022. [407]

Newey, W., and McFadden, D. (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics* (Vol. 36), eds. R. F. Engle and D. L. McFadden as Amsterdam: North Holland, pp. 2111–2234. [405,408]

Phillips, P. (2012), "Folklore Theorems, Implicit Maps, and Indirect Inference," *Econometrica*, 80, 425–454. [409]

Ramberg, J., and Schmeiser, B. (1975), "An Approximate Method for Generating Asymmetric Random Variables," *Communications of the ACM*, 17, 78–82. [410,416]

Ramsey, J., and Montenegro, A. (1992), "Identification and Estimation of Non-Invertible Non-Gaussian MA(q) processes," *Journal of Econometrics*, 54, 301–320. [406]

Rothenberg, T. (1971), "Identification in Parametric Models," *Econometrica*, 39, 577–591. [405]

Sargan, D., and Bhargava, A. (1983), "Maximum Likelihood Estimation of Regression Models With First Order Moving Average Errors When the Root Lies on the Unit Circle," *Econometrica*, 51, 799–820. [405]

Sargan, J. D. (1983), "Identification and Lack of Identification," *Econometrica*, 51, 1605–1633. [405]

Tauchen, G. (1998), "The Objective Function of Simulation Estimators Near the Boundary of the Unstable Region of the Parameter Space," *Review of Economics and Statistics*, 80, 389–398. [410]

Tugnait, J. (1986), "Identification of Non-Minimum Phase Linear Stochastic Systems," *Automatica*, 22, 457–464. [406]

——— (1995), "Parameter Estimation for Noncausal ARMA Models of Non-Gaussian Signals via Cumulant Matching," *IEEE Transactions on Signal Processing*, 43, 886–894. [407,416]