Minimum Distance Estimation of Possibly Non-Invertible Moving Average Models

Nikolay Gospodinov* Serena Ng †

October 25, 2013

Abstract

This paper 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 has classical (root-T and asymptotic normal) properties when the moving average 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 that distinguish them from the existing work in the literature. First, identification now requires information from the second and higher order moments of the data. Thus, in addition to an autoregressive model, new auxiliary regressions need to be 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 non-invertibility in the Fama-French portfolio returns.

JEL Classification: C13, C15, C22

Keywords: GMM; Simulation-based estimation; Non-invertibility; Identification; Non-Gaussian errors; Generalized lambda distribution

---

*Federal Reserve Bank of Atlanta, Research Department, 1000 Peachtree Street N.E., Atlanta, GA 30309–4470. Email: nikolay.gospodinov@atl.frb.org

†Columbia University, Department of Economics, 420 West 118th Street, Mail Code 3308, New York, NY 10027. Email: serena.ng@columbia.edu

We would like to 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.

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 non-invertible moving average processes are observationally equivalent up to the second moments. Second, invertibility restricts all roots of the moving average polynomial to be less than or equal to one. This upper bound renders estimators with non-normal asymptotic distributions when some roots are on or near the unit circle. Existing estimators treat invertible and non-invertible processes separately, requiring the researcher to take a stand on the parameter space of interest. While the estimators are super-consistent under the null hypothesis of a moving average unit root, their distributions are not asymptotically pivotal. To our knowledge, no estimator of the MA model exists that achieves identification without imposing invertibility and yet enables classical inference over the whole parameter space.

Both invertible and non-invertible 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 non-invertibility can arise if the variance of the stationary component is large. While non-invertible 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.[1] 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 (see Fernández-Villaverde et al. (2007) for an example regarding the permanent income model). Hansen and Sargent (1991), Lippi and Reichlin (1993), Fernández-Villaverde et al. (2007), among others, emphasize the need to verify invertibility because it affects how we interpret what can be recovered from the data.

Indeed, it is necessary in many science and engineering applications to admit parameter values in the non-invertible range.[2] A key finding in these studies is that higher order cumulants are necessary for identification of non-invertible models, implying that the assumption of Gaussian errors must be abandoned. Lii and Rosenblatt (1992) approximate the non-Gaussian likelihood of non-invertible MA models by truncating the representation of the innovations in terms of the

[1]Invertibility can also help to identify structural models. For example, Komunjer and Ng (2011) use invertibility to narrow the class of equivalent DSGE models.

[2]For example, in seismology, an accurate model of the seismic source wavelet, in the form of a moving average filter, is necessary to recover the earth's reflectivity sequence. The fact that seismic data typically exhibit non-Gaussian features suggests the need for a wavelet (moving average polynomial) which is non-invertible. Similarly, in communication analysis, an accurate modeling of the communication channel by a possibly non-invertible moving average process is required to back out the underlying message from the observed distorted message.

observables. Huang and Pawitan (2000) propose least absolute deviations (LAD) estimation using a Laplace likelihood. This quasi maximum likelihood estimator does not require the errors to be Laplace distributed, but they need to have heavy tails. Andrews et al. (2006, 2007) consider LAD and rank-based estimation of all-pass models.[3] Meitz and Saikkonen (2011) develop maximum likelihood estimation of non-invertible ARMA models with ARCH errors. However, there exist no likelihood based estimators that have classical properties while admitting a moving-average unit root in the parameter space.

This paper 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 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 non-invertible moving-average 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 a generalized method of moments (GMM) estimator that is root-T consistent and asymptotically normal without restricting the moving average roots to be strictly inside the unit circle.

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 et al. (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. 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 paper proceeds as follows. Section 2 highlights two identification problems that arise in

[3] All-pass models are special non-causal and/or non-invertible autoregressive and moving average (ARMA) models in which the roots of the autoregressive polynomial are reciprocals of the roots of the moving average polynomials.

moving average 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 moving average 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 $L^p y_t = y_{t-p}$, $\alpha(L) = 1 - \alpha_1 L - \ldots - \alpha_p L^p$ have no common roots with $\theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q$. 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 moving average 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 moving average models without prior knowledge about invertibility. The distinction between invertible and non-invertible processes is best illustrated by considering the MA(1) model defined by

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

with $e_t \sim iid(0, \sigma^2)$. 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 a moving-average 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 non-invertibility boundary; see Sargan and Bhargava (1983), Anderson and Takemura (1986), Davis and Dunsmuir (1996), Gospodinov (2002), 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 non-invertible 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 towards 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 in 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 Non-Invertible 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 non-linear equations. For non-linear models, a sufficient condition is that the Jacobian matrix of the first partial derivatives is full column rank. However, local identification is still possible if the rank condition fails by exploiting restrictions on the higher order derivatives.[4] To obtain results for global identification, Rothenberg (1971, Theorem 7) imposed additional conditions to ensure that the optimization problem is well behaved. In a semi-parametric 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)).[5]

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), $\gamma = (\alpha_1, ..., \alpha_p, \theta_1, \ldots, \theta_q, \sigma^2)'$. 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 $g(\gamma) : \gamma \to g(\gamma)$ be a mapping from γ to $g(\cdot)$ and $G(\gamma) = \partial g(\gamma)/\partial \gamma'$ with $G_0 \equiv G(\gamma_0)$. Then, γ_0 is globally identified 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 $g(\gamma_1) = g(\gamma_2)$, i.e., $g(\cdot)$ is not injective. Subsection 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 non-invertible models can be identified if $g(\gamma)$ is allowed to include higher order moments/cumulants. Subsection 3.2 generalizes 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 is 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 identification problem using only second moments of y_t:

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

[4]See Sargan (1983), Durlauf and Blume (2008), and Dovonon and Renault (2011) for more examples.

[5]Komunjer (2012) shows that global identification from moment restrctions is possible even when the derivative matrix has a deficient rank, provided that this happens only over sufficiently small regions in the parameter space.

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

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

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 non-identification and non-normality 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 non-zero 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 $\varepsilon_t \sim iid(0, 1)$ with $\kappa_3 = E(\varepsilon_t^3)$. Assume that $\theta \neq 0$, $\kappa_3 \neq 0$ and $E|\varepsilon_t|^6 < \infty$. 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 over-identified 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 MA(q) and ARMA(p, q) Models

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 non-linear functions of the model parameters and verifying global and local identification is more challenging. We capitalize on an insight from the statistical engineering literature and augment the original parameters of interest with some nonlinear transformations of these parameters such that the augmented parameter vector is a solution to a linear system of equations of second and higher order cumulants. As a result, the identifiability of the MA parameters boils down to a full column rank requirement on a matrix consisting of population cumulants that can be consistently estimated from the data.

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

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

If $\tau_1 = \tau_2 = ... = \tau_\ell = \tau$, $c_\ell(\tau) = c_\ell(\tau, ..., \tau)$ is known as the diagonal slice of the ℓ-th order cumulant of y_t. If $y_t = h(L)e_t$ and e_t are *iid* with finite ℓ-th order cumulant η_ℓ (noting that $\eta_2 = \sigma^2$). Then,

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

The cumulants η_ℓ ($\ell \geq 3$) measure the distance of the stochastic process from Gaussianity.

Higher order cumulants are useful for identification of possibly non-invertible models because the Fourier transform of $c_\ell(\tau_1, \tau_2, ..., \tau_{\ell-1})$ is the ℓ-th order polyspectrum

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

Recovery of phase information necessarily requires that e_t has non-Gaussian features. Provided that η_ℓ exists and is non-zero for $\ell \geq 3$, one can recover the phase function from any ℓ-th order spectrum, see Lii and Rosenblatt (1982, Lemma 1), Giannakis and Swami (1992), Giannakis and Mendel (1989), Mendel (1991), Tugnait (1986), Ramsey and Montenegro (1992).

Establishing that the parameters are identified (that is, can be expressed in terms of the population cumulants) is non-trivial because the mapping from the cumulants to the parameters is non-linear. One approach in engineering literature is to use the spectrum $S_2(z) = \sigma^2 H(z)H(z^{-1})$ to substitute out $H(z^{-1})$ in the polyspectra corresponding to $c_\ell(\tau_1, \tau_2, ..., \tau_{\ell-1})$ for a particular choice of $\tau_1, \ldots, \tau_{\ell-1}$.[6] This generates an identity in time domain that links the population second and higher order cumulants to the parameters of the model. For the MA(q) model, the diagonal slice of the third order cumulants implies the following relation between the population cumulants and the parameters $(\theta_1, \ldots, \theta_q, \widetilde{\eta}_3)$ where $\widetilde{\eta}_3 = \eta_3/\sigma^2$:

$$\sum_{j=1}^{q} \theta_j c_3(\tau - j) - \widetilde{\eta}_3 \sum_{j=0}^{q} \theta_j^2 c_2(\tau - j) + c_3(\tau) = 0, \quad -q \leq \tau \leq 2q, \tag{7}$$

To establish identification, define $\theta = (\theta_1, \ldots, \theta_q)'$,

$$\beta(\theta, \widetilde{\eta}_3) = (\theta_1, \ldots, \theta_q, \widetilde{\eta}_3, \widetilde{\eta}_3\theta_1^2, \ldots, \widetilde{\eta}_3\theta_q^2)',$$

$$b = \begin{bmatrix} -c_3(-q) & -c_3(-q+1) & \cdots & -c_3(0) & -c_3(1) & \cdots & -c_3(q-1) & -c_3(q) & 0 & 0 & \cdots & 0 \end{bmatrix}'.$$

Also let

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

where

$$B = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ c_3(-q) & 0 & \ldots & 0 \\ c_3(-q+1) & c_3(-q) & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ c_3(q-1) & c_3(q-2) & \ldots & c_3(0) \end{bmatrix}, \quad E = \begin{bmatrix} 0 & -c_2(q) & -c_2(q-1) & \cdots & -c_2(1) \\ 0 & 0 & -c_2(q) & \cdots & -c_2(2) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & -c_2(q) \end{bmatrix},$$

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

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

[6] For $\ell = 3$, $S_3(z) = \eta_3 H(z^{-1})H(z) * H(z)$, where $*$ is the complex convolution operator.

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

The identities defined by (7) can be expressed as an overidentified system of $3q+1$ equations in $2q+1$ unknowns:

$$A\beta(\theta, \widetilde{\eta}_3) = b. \tag{8}$$

Identification of the MA coefficients $\theta = (\theta_1, ..., \theta_q)'$ is now reduced to two problems: identification of $\beta(\theta, \widetilde{\eta}_3)$, and identification of $(\theta, \widetilde{\eta}_3)$ from β. As the derivative matrix of β with respect to θ has rank $2q+1$, the first problem reduces to the verification of the column rank of the matrix A.

Lemma 2 *Consider the MA(q) process $y_t = e_t + \theta_1 e_{t-1} + ... + \theta_q e_{t-q}$. Let $c_\ell(\tau)$ denote the diagonal slice of the ℓ-th order cumulant of y_t and assume that $c_2(q) \neq 0$, $c_3(q) \neq 0$ and $E|e_t|^6 < \infty$. Then, (i) the matrix A has full column rank $2q+1$, and (ii) $\theta = (\theta_1, ..., \theta_q)'$ is identifiable.*

The rank of the $(3q+1) \times (2q+1)$ matrix A is the sum of the column rank of the sub-matrix consisting of B and D, and the rank of the sub-matrix consisting of C and E. The rank of the first sub-block 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 non-zero. Lemma 2 implies that $\beta(\theta, \widetilde{\eta})$ is identified. Identifiability of the parameters in MA(q) models follows from identifiability of β. Primitive conditions for ensuring this identifiability are $\theta_1 \neq 0, ..., \theta_q \neq 0$ and $\eta_3 \neq 0$.

We can further incorporate information about the fourth order cumulants of y_t.[7] This is particularly useful when the signal from the third order moments is weak or zero (i.e., the error distribution is (near-) symmetric). The equations based on the diagonal slice of c_4 are

$$\sum_{i=1}^{q} \theta_i c_4(\tau - i) - \widetilde{\eta}_4 \sum_{i=0}^{q} \theta_i^3 c_2(\tau - i) = -c_4(\tau),$$

where $\widetilde{\eta}_4 = \eta_4/\sigma^2$, $c_4(-1) = c_4(1,0,0) = E(y_t^3 y_{t+1}) - 3E(y_t^2)E(y_t y_{t+1})$, $c_4(0) = E(y_t^4) - 3[E(y_t^2)]^2$ and $c_4(1) = E(y_t y_{t+1}^3) - 3E(y_t^2)E(y_t y_{t+1})$. The augmented vector of parameters $(\theta_1, ..., \theta_q, \widetilde{\eta}_3, \widetilde{\eta}_3\theta_1^2, ..., \widetilde{\eta}_3\theta_q^2, \widetilde{\eta}_4, \widetilde{\eta}_4\theta_1^3, ..., \widetilde{\eta}_4\theta_q^3)'$ is identified provided that $\eta_3, \eta_4 \neq 0$.

The $A\beta(\theta, \widetilde{\eta}_3) = b$ approach requires that q is finite and hence does not work for ARMA(p, q) models. The identification of ARMA(p, q) models raises some additional difficulties due to (i) the possibility of cancelling roots in the AR and MA polynomials which is itself an identification problem and (ii) the infinite order MA structure generated by the model. The literature on identification

[7] The fourth order cumulant is defined as $c_4(\tau_1, \tau_2, \tau_3) = E(y_t y_{t+\tau_1} y_{t+\tau_2} y_{t+\tau_3}) - c_2(\tau_1)c_2(\tau_2 - \tau_3) - c_2(\tau_2)c_2(\tau_3 - \tau_1) - c_2(\tau_3)c_2(\tau_1 - \tau_2)$. Non-diagonal slices of the fourth-order cumulants were considered in Friedlander and Porat (1990) and Na et al. (1995).

of possibly non-minimum phase ARMA(p, q) models is quite small and typically proceeds by assuming that the ARMA model has no common factors (hence, it is irreducible). The following lemma provides sufficient conditions for identifiability of the parameters of ARMA(p, q) models.

Lemma 3 *Assume that the ARMA(p, q) process* $(1-\alpha_1 L-\ldots-\alpha_p L^p)y_t=(1+\theta_1 L+\ldots+\theta_q L^q)e_t$*, where* e_t *is a zero-mean iid process, is irreducible and satisfies* $\sum_{i=0}^{p}\alpha_i z^i \neq 0$ *for* $|z|=1$*. Let* $c_\ell(\tau)$ *denote the diagonal slice of the* ℓ*-th order cumulant of the MA(p+q) process* $(1-\alpha_1 L-\ldots-\alpha_p L^p)(1+\theta_1 L+\ldots+\theta_q L^q)e_t$ *and assume that* $c_2(p+q)\neq 0$, $c_3(p+q)\neq 0$ *and* $E|e_t|^6<\infty$*. Then, the parameter vector* $(\alpha_1,\ldots,\alpha_p,\theta_1,\ldots,\theta_q)'$ *is identifiable from the second and third cumulants of the process.*

The proof of Lemma 3 exploits the results available for identification of moving average models by introducing an auxiliary process with an MA(p+q) component. Tugnait (1995) used information in the non-diagonal slices to isolate the smallest number of third and higher order cumulants that are sufficient for identification of ARMA parameters (but not σ^2).

The representation $A\beta(\theta,\widetilde{\eta}_3)=b$ 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 not be efficient. For example, it may be possible to replace some equations in the system by Yule-Walker equations which would help identify the autoregressive parameters. Furthermore, (5) implies that for an MA(q) process, $c_3(q,k)=\eta_3\theta_q\theta_k$ and $c_3(q,0)=\eta_3\theta_q$. It immediately follows that $\theta_k=\frac{c_3(q,k)}{c_3(q,0)}$. This is the so-called $C(q,k)$ formula. Hence, 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 $\eta_3\neq 0$, which is smaller than the number of equations in the $A\beta(\theta,\widetilde{\eta}_3)=b$ system. See Mendel (1991) for a survey of the methods used in the engineering literature. The point to highlight is that once non-Gaussian features are allowed, identification of non-invertible models is possible from the higher order cumulants of the data.

4 GMM Estimation

The results in Section 3 suggest two estimators within the method of moments framework. In particular, Lemma 2 implies $3q+1$ orthogonality conditions for identifying $\widetilde{\eta}_3$ and $\theta_1,\ldots\theta_q$ of the MA(q) model. In the MA(1) case, the population orthogonality condition has the form

$$g(\theta,\widetilde{\eta}_3)=\begin{bmatrix} 0 & -c_2(1) & 0 \\ c_3(-1) & -c_2(0) & -c_2(1) \\ c_3(0) & -c_2(1) & -c_2(0) \\ c_3(1) & 0 & -c_2(1) \end{bmatrix}\begin{bmatrix} \theta \\ \widetilde{\eta}_3 \\ \widetilde{\eta}_3\theta^2 \end{bmatrix}-\begin{bmatrix} -c_3(-1) \\ -c_3(0) \\ -c_3(1) \\ 0 \end{bmatrix} \tag{9}$$

which equals zero at the true values of θ and $\widetilde{\eta}_3$. As the equations in (9) do not separately identify σ^2 and η_3 (or κ_3), additional conditions from the autocovariances need to be appended. It is thus more convenient to estimate $\gamma = (\theta, \sigma^2, \kappa_3)'$ from the moment conditions implied by Lemma 1:

$$g_{GMM}(\gamma) = \begin{pmatrix} E(y_t y_{t-1}) \\ E(y_t^2) \\ E(y_t^2 y_{t-1}) \\ E(y_t^3) \\ E(y_t y_{t-1}^2) \end{pmatrix} - \begin{pmatrix} \theta\sigma^2 \\ (1+\theta^2)\sigma^2 \\ \theta^2\sigma^3\kappa_3 \\ (1+\theta^3)\sigma^3\kappa_3 \\ \theta\sigma^3\kappa_3 \end{pmatrix} = \begin{pmatrix} c_2(1) \\ c_2(0) \\ c_3(1) \\ c_3(0) \\ c_3(-1) \end{pmatrix} - \begin{pmatrix} \theta\sigma^2 \\ (1+\theta^2)\sigma^2 \\ \theta^2\eta_3 \\ (1+\theta^3)\eta_3 \\ \theta\eta_3 \end{pmatrix} \tag{10}$$

since $E(y_t y_{t-1}) = c_2(1)$, $E(y_t^2) = c_2(0)$, $E(y_t^2 y_{t-1}) = c_3(1)$, $E(y_t^3) = c_3(0)$, $E(y_t y_{t-1}^2) = c_3(-1)$ and $\eta_3 = \sigma^3\kappa_3$. Thus, the moment conditions in (9) are particular linear combinations of the moment conditions in (10). 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.

Given data $\mathbf{y} \equiv (y_1, \ldots, y_T)'$, one can construct $\widehat{g}_{MMM}(\gamma)$, the sample analog of $g_{GMM}(\gamma)$ defined in (10).[8] Let $\widehat{\Omega}_{GMM}$ denote a consistent estimate of $\Omega_{GMM} = \lim_{T\to\infty} \text{Var}(g_{GMM}(\gamma_0))$, where Ω_{GMM} is positive definite. Then, the optimal GMM estimator of $\gamma = (\theta, \sigma^2, \kappa_3)'$ is defined as

$$\widehat{\gamma}_{GMM} = \arg\min_\gamma \widehat{g}_{GMM}(\gamma)'\widehat{\Omega}_{GMM}^{-1}\widehat{g}_{GMM}(\gamma). \tag{11}$$

The derivative matrix of $g_{GMM}(\gamma)$ with respect to γ, $G_{GMM}(\gamma)$, is of full column rank everywhere in Γ (even at $|\theta| = 1$) which is sufficient for γ_0 to be a unique solution to the system of non-linear equations characterized by

$$G_{GMM}(\gamma)'\Omega_{GMM}^{-1}g_{GMM}(\gamma) = 0.$$

The full rank condition, in the neighborhood of γ_0, is also necessary for the estimator to be asymptotically normal. As a result, this GMM estimator is root-T consistent and asymptotic normal.

Proposition 1 *Consider the MA(1) model and suppose that the conditions in Lemma 1 hold. In addition, assume that γ_0 is in the interior of the compact parameter space Γ, $\sqrt{T}(\widehat{g}_{GMM}(\gamma) - g_{GMM}(\gamma_0)) \xrightarrow{d} N(0, \Omega_{GMM})$ and $\widehat{G}_{GMM}(\gamma)$ converges uniformly to $G_{GMM}(\gamma)$ over $\gamma \in \Gamma$. Then,*

$$\sqrt{T}(\widehat{\gamma}_{GMM} - \gamma_0) \xrightarrow{d} N\Big(0, \big(G_{GMM}(\gamma_0)'\Omega_{GMM}^{-1}G_{GMM}(\gamma_0)\big)^{-1}\Big).$$

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

[8] Unreported numerical results revealed that the estimator based on the moment conditions (10) possesses substantially better finite-sample properties than the estimator based on (9).

conditions should be selected in an optimal fashion. 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.

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 generalized lambda distribution (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.

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 (10) as moment conditions. The second is the infeasible GMM estimator based on (10) 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 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 non-invertible 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 non-invertible 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.

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 non-trivial 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 et al. (1999) and 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 et al. (1993), Michaelides and Ng (2000), Ghysels et al. (2003), 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 et al. (1993) find 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 et al. (2003) also evaluate the properties of simulation-based estimators with σ^2 assumed known. Czellar and Zivot (2008) report 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 generalized lambda distribution $\Lambda(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ 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

$$\Lambda(u)^{-1} = \lambda_1 + [U^{\lambda_3} + (1-U)^{\lambda_4}]/\lambda_2, \tag{12}$$

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 (12) 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

Let $\mathbf{y}^S(\gamma) = (y_1^S, \ldots, y_T^S, ..., y_{TS}^S)'$ be data of length TS ($S \geq 1$), simulated for a candidate value of γ. This usually requires drawing errors from a known distribution, and the parameters of this distribution are ancillary for γ. Let

$$\widehat{g}_{SMM}(\gamma) = \frac{1}{T}\sum_{t=1}^{T} m_{SMM,t}(\mathbf{y}) - \frac{1}{TS}\sum_{t=1}^{TS} m_{SMM,t}^S(\mathbf{y}^S(\gamma)),$$

where $m_{SMM,t}(\mathbf{y})$ and $m_{SMM,t}^S(\mathbf{y}^S(\gamma))$ denote vectors of moments based on actual and simulated data, respectively, at time t. Define the augmented parameter vector of the MA(1) model as $\gamma = (\theta, \sigma^2, \lambda_3, \lambda_4)'$. The simulated method of moments (SMM) estimator is

$$\widehat{\gamma}_{SMM} \equiv \arg\min_\gamma \widehat{g}_{SMM}(\gamma)'\widehat{\Omega}_{SMM}^{-1}\widehat{g}_{SMM}(\gamma), \tag{13}$$

where $\widehat{\Omega}_{SMM}$ is a consistent estimate of the long-run asymptotic variance of $m_{SMM,t}(\mathbf{y})$. Identification requires that $g_{SMM}(\cdot)$ be injective in the sense of Definition 1.[9]

[9] It would seem tempting to estimate λ_3 and λ_4 separately from $(\theta, \sigma^2)'$, such as using the sample skewness and kurtosis of the residuals of a long autoregression. But as discussed in Ramsey and Montenegro (1992), the OLS

It remains to define $m_{SMM,t}$. 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 errors from estimation of an AR(p) model

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

The moment conditions given by

$$m_{SMM,t}(\mathbf{y}) = \begin{pmatrix} y_t y_{t-1} & y_t^2 & y_t^2 y_{t-1} & y_t^3 & y_t y_{t-1}^2 & y_t^3 y_{t-1} & y_t y_{t-1}^3 & y_t^2 y_{t-1}^2 & y_t^4 & \widehat{\epsilon}_t^3 & \widehat{\epsilon}_t^4 \end{pmatrix}' \quad (14)$$

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 $\widehat{\gamma}_{SMM}$ we need some additional notation and regularity conditions. Let $m_{SMM}(\mathbf{y}) = E(m_{SMM,t}(\mathbf{y}))$, F_e denote the true distribution of the structural model errors and Λ_* be the class of generalized lambda distributions.

Proposition 2 *Suppose that in addition to the assumptions in Lemma 1, we have* $F_e \in \Lambda_*$, $E|e_t|^8 < \infty$, $sup_{\gamma\in\Gamma}\left|\widehat{G}_{SMM}(\gamma) - G_{SMM}(\gamma)\right| \xrightarrow{p} 0$, γ_0 *is in the interior of the compact parameter space* Γ, *and* $\sqrt{T}(\widehat{m}_{SMM} - m_{SMM}) \xrightarrow{d} N(0, \Omega_{SMM})$. *Then,*

$$\sqrt{T}(\widehat{\gamma}_{SMM} - \gamma_0) \xrightarrow{d} N\left(0, \left(1 + \frac{1}{S}\right)\left(G_{SMM}(\gamma_0)'\Omega_{SMM}^{-1} G_{SMM}(\gamma_0)\right)^{-1}\right) \equiv N\Big(0, \text{Avar}(\widehat{\gamma}_{SMM})\Big).$$

Consistency follows from identifiability of γ and the moment conditions that exploit information in 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 non-standard 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 $\widehat{\gamma}_{SMM}$ has classical properties provided that κ_3 and κ_4 enable identification.

Consistent estimation of the asymptotic variance of $\widehat{\gamma}_{SMM}$ can proceed by substituting a consistent estimator of Ω_{SMM} and evaluating the Jacobian $G_{SMM}(\widehat{\gamma}_{T,S})$ numerically. The computed

residuals do not converge in the limit to the true errors when $\theta(L)$ is non-invertible, rendering their sample higher moments also asymptotically biased.

standard errors can then be used for testing hypotheses and constructing confidence intervals.[10] Alternatively, inference on the MA parameter of interest, θ, can be conducted by constructing confidence intervals based on inversion of the distance metric test without an explicit computation of the variance matrix $\text{Avar}(\widehat{\gamma}_{SMM})$.

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. Our simulated minimum distance (SMD) is defined as

$$\widehat{\gamma}_{SMD} \equiv \arg\min_{\gamma}(\widehat{\psi}_{SMD} - \widetilde{\psi}^{S}_{SMD}(\gamma))'\widehat{\Omega}^{-1}_{SMD}(\widehat{\psi}_{SMD} - \widetilde{\psi}^{S}_{SMD}(\gamma)), \tag{15}$$

where $\widehat{\psi}_{SMD} = \arg\min_{\psi} Q_T(\psi;\mathbf{y})$ and $\widetilde{\psi}^{S}_{SMD}(\gamma) = \arg\min_{\psi} Q_T(\psi;\mathbf{y}^S(\gamma))$ are the auxiliary parameters estimated from actual and simulated data, $Q_T(\cdot)$ denotes the objective function of the auxiliary model and $\widehat{\Omega}_{SMD}$ is a consistent estimate of the asymptotic variance of $\widehat{\psi}_{SMD}$.

Our SMD estimator is in the spirit of the indirect inference estimation of Gourieroux et al. (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. For estimation of ARMA models, we consider least squares estimation of the auxiliary regressions

$$y_t = \pi_0 + \pi_1 y_{t-1} + \ldots + \pi_p y_{t-p} + \sigma\epsilon_t, \tag{16a}$$

$$y_t^2 = c_0 + c_{1,1}y_{t-1} + \ldots + c_{1,r}y_{t-r} + c_{2,1}y^2_{t-1} + \ldots + c_{2,r}y^2_{t-r} + v_t \tag{16b}$$

with an appropriate choice of p and r. Equation (16a) 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 (16a) with the regression defined in (16b). 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_{SMD} are

[10] 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 semi-parametric indirect inference estimator of Dridi et al. (2007) in partially misspecified structural models which requires an adjustment in the asymptotic variance of the estimator.

based on conditional moments. Equation (16b) 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 OLS residuals in (16a). The auxiliary parameter vector based on the data is

$$\widehat{\psi}_{SMD} = (\widehat{\pi}_0, \widehat{\pi}_1, ..., \widehat{\pi}_p, \widehat{c}_0, \widehat{c}_{1,1}, ..., \widehat{c}_{1,r}, \widehat{c}_{2,1}, ..., \widehat{c}_{2,r}, \widehat{\kappa}_3, \widehat{\kappa}_4)'. \tag{16c}$$

The parameter vector $\psi^S_{SMD}(\gamma)$ is analogously defined, except that the auxiliary regressions are estimated with data simulated for a candidate value of γ. The optimal SMD estimator shares the same asymptotic properties as the SMM estimator in Proposition 2.

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 $S = 20$ 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 (16b).

As is true of all non-linear estimation problems, the numerical optimization problem must take into account the possibility of local minima which arise 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 $\sqrt{\widehat{\sigma}^2_y/(1+\theta^2)}$ 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): First, we study the finite-sample behavior of the proposed SMM and SMD estimators in invertible and non-invertible MA(1) models with data generated from

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

where $\varepsilon_t \sim iid(0, 1)$ is drawn from a GLD with zero excess kurtosis and a skewness parameter 0.85 with (i) $\sigma_t = \sigma = 1$ or (ii) $\sigma_t = 0.7 + 0.3e^2_{t-1}$ (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, i.e., 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 (for example, $\theta_0 = 2$). While there is a positive probability that the SMM estimator will converge to $1/\theta$ instead of θ (especially when θ is in the non-invertible 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/non-invertible 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.[11] 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

$$y_t = \alpha y_{t-1} + e_t + \theta e_{t-1}, \tag{17}$$

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 of the model that give rise to a causal process with a non-invertible MA component. The first parameterization is $\alpha = 0.5$ and $\theta = -1.5$. The second

[11] In evaluating the objective function, the values of the lambda parameters in the generalized lambda distribution are set equal to their true values.

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 heteroskedastic. 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 (17) 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

Non-invertibility 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 $x_t = e_t + \omega e_{t-1}$ (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 a moving average unit root when $\omega = -0.5$ and $h(L)$ is non-invertible in the past whenever $\omega < -0.5$.[12]

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 5 portfolios formed on size (market equity) and 5 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

[12] If the moving average polynomial $\omega(L)$ is of infinite order, as it would be the case for causal autoregressive processes, 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.

of the returns exhibit some form of non-Gaussianity, which is necessary for identifying possible non-invertible 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 by a non-invertible 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 non-invertibility in stock returns.

6 Conclusions

This paper proposes generalized and simulation-based method of moments estimation of possibly non-invertible 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 non-standard asymptotics of the Gaussian maximum likelihood estimator. As a consequence, the proposed GMM estimator for the MA(1) model is 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.

To accommodate more general models which require many higher order moments that are analytically intractable or cannot be precisely estimated in finite samples, we develop two simulation estimators that incorporate information from the higher order cumulants of the data. The efficiency of the estimators is controlled by the ability of the auxiliary moments in approximating the true data generating process. Our proposed estimators use an error simulator with a flexible functional form that encompasses a large class of distributions with non-Gaussian features. Particular attention is paid to the accurate estimation of the shape parameters of the error distribution which play a critical role in identifying the structural parameters.

A 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

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

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. 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 $\beta(\theta, \eta)$ can be obtained as a unique solution to the system of equations (8). Since the derivative matrix of $\beta(\theta, \widetilde{\eta}_3)$ given by

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 1 \\ 2\widetilde{\eta}_3\theta_1 & 0 & \cdots & 0 & \theta_1^2 \\ 0 & 2\widetilde{\eta}_3\theta_2 & \cdots & 0 & \theta_2^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 2\widetilde{\eta}_3\theta_q & \theta_q^2 \end{bmatrix}$$

is of full column rank, the parameter vector of interest $(\theta_1, ..., \theta_q)'$ is identifiable.

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 $z_t = \alpha_2(L)\theta_1(L)e_t$ and $z_t = \alpha_1(L)\theta_2(L)e_t$, where $z_t = a_1(L)a_2(L)y_t$. Let $\gamma_1 = (\alpha_{1,1}, ..., \alpha_{1,p}, \theta_{1,1}, ..., \theta_{1,q})'$ and $\gamma_2 = (\alpha_{2,1}, ..., \alpha_{2,p}, \theta_{2,1}, ..., \theta_{2,q})'$. Note that these are two MA(p+q) processes for z_t, $z_t = \Theta_1(L; \gamma_1, \gamma_2)$ and $z_t = \Theta_2(L; \gamma_1, \gamma_2)$, where $\Theta_1(\cdot)$ and $\Theta_2(\cdot)$ denote MA polynomials of order $p+q$.

As in Lemma 2, we can write

$$
\begin{aligned}
A\beta_1(\gamma_1, \eta) &= b \\
A\beta_2(\gamma_2, \eta) &= b,
\end{aligned}
\tag{18}
$$

where A and b are functions of second and third cumulants of z_t. But from Lemma 2, there is a unique solution to the system of equations $A\beta(\gamma, \eta) = b$. Hence, there is a one-to-one mapping between (A, b) and $\beta(\gamma, \eta)$ and the two ARMA models are identical in the sense that $\gamma_1 = \gamma_2$. Therefore, $\gamma = (\alpha_1, ..., \alpha_p, \theta_1, ..., \theta_q)'$ is identifiable from the second and third cumulants used in constructing A and b, provided that $c_2(p+q) \neq 0$ and $c_3(p+q) \neq 0$.

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)):

$$
\begin{aligned}
\kappa_3 &= \frac{C - 3AB + 2A^3}{\lambda_2^3}, \\
\kappa_4 &= \frac{D - 4AC + 6A^2B - 3A^4}{\lambda_2^4},
\end{aligned}
$$

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.

References

Anderson, T. W. and Takemura, A. 1986, Why Do Noninvertible Estimated Moving Average Models Occur?, *Journal of Time Series Analysis* **7**(4), 235–254.

Andrews, B., Davis, R. and Breidt, F. 2007, Rank-Based Estimation of All-Pass Time Series Models, *Annals of Statistics* **35**, 844–869.

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.

Brockwell, P. J. and Davies, R. A. 1991, *Time Series Theory and Methods*, 2nd edn, Springer-Verlag, New York.

Czellar, V. and Zivot, E. 2008, Improved Small Sample Inference for Efficient Method of Moments and Indirect Inference Estimators, University of Washington.

Davis, R. 2010, All-Pass Procssess with Applications to Finance, *Plenary Talk at the 7th International Iranian Workshop on Stocastic Processes*.

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.

Davis, R. and Song, L. 2011, Unit Roots in Moving Averages Beyond First Order, *Annals of Statistics* **39**(6), 3062–3091.

Dovonon, P. and Renault, E. 2011, Testing for Common GARCH Factors, MPRA paper 40244.

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

Duffie, D. and Singleton, K. 1993, Simulated Moments Estimation of Markov Models of Asset Prices, *Econometrica* **61**, 929–952.

Durlauf, S. and Blume, L. (eds) 2008, *Identification*, Vol. Second Edition, Palgrave Macmillan.

Fernández-Villaverde, F., Rubio-Ramírez, J., Sargent, T. and Watson, M. 2007, A,B,Cs and (D)'s for Understanding VARs, *American Economic Review* **97:3**, 1021–1026.

Fisher, F. M. 1961, Identifiability Criteria in Non-Linear Systems, *Econometrica* **29**, 574–590.

Fisher, F. M. 1965, Identifiability Criteria in Non-Linear Systems: A Further Note, *Econometrica* **33**, 197–206.

Friedlander, B. and Porat, B. 1990, Asymptotically Optimal Estimation of MA and ARMA Parameters of Non-Gaussian Processes of Higher-Order Momentes, *IEEE Transactions on Automatic Control* **35**(1), 27–35.

Gallant, R. and Tauchen, G. 1996, Which Moments to Match, *Econometric Theory* **12**, 657–681.

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

Giannakis, G. and Mendel, J. 1989, Identification of Nonminimum Phase Systems Using Higher Order Statistics, *IEEE Transations, Acoustics, Speech, and Signal Processing* **37**(3), 360–377.

Giannakis, G. and Swami, A. 1992, Identifiability of General ARMA Processes Using Linear Cumulant-Based Estimators, *Automatica* **28**(4), 771–779.

Gospodinov, N. 2002, Bootstrap Based Inference in Models with a Nearly Noninvertible Moving Average Component, *Journal of Business and Economic Statistics* **20**, 254–268.

Gourieroux, C., Monfort, A. and Renault, E. 1993, Indirect Inference, *Journal of Applied Econometrics* **85**, 85–118.

Gourieroux, C., Renault, E. and Touzi, N. 1999, Calibration by Simulation for Small Sample Bias Correction,, *in* R. Mariano, T. Schuermann and M. Weeks (eds), *Simulation-based Inference in Econometrics: Methods and Applications*, Cambridge University Press.

Hansen, L. and Sargent, T. 1991, Two Difficulties in Interpreting Vector Autoressions, *in* L. P. Hansen and T. J. Sargent (eds), *Rational Expectations Econometrics*, Westview, London, pp. 77–119.

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

Komunjer, I. 2012, Global Identification in Nonlinear Models with Moment Restrictions, *Econometric Theory* **28**, 719–729.

Komunjer, I. and Ng, S. 2011, Dynamic Identification of Dynamic Stochastic General Equilibrium Models, *Econometrica* **79:6**, 1995–2032.

Lii, K. and Rosenblatt, M. 1982, Deconvolution and Estimation of Transfer Function Phase Coefficients for Non-Gaussian Linear Processes, *Annals of Statistics* **10**, 1195–1208.

Lii, K. and Rosenblatt, M. 1992, An Approximate Maximum Likelihood Estiation of Non-Gaussian Non-Minimum Phase Moving Average Processes, *Journal of Multivariate Analysis* **43**, 272–299.

Lippi, M. and Reichlin, L. 1993, The Dynamic Effects of Aggregate Demand and Supply Disturbances: Comment, *American Economic Review* **83**, 644–52.

Meitz, M. and Saikkonen, P. 2011, Maximum Likelihood Estimation of a Non-Invertible ARMA Model with Autoregressive Conditional Heteroskedasticity, mimeo, University of Helsinki.

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**(3), 278–305.

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* **96:2**, 231–266.

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**(5), 2018–2022.

Newey, W. and McFadden, D. 1994, Large Sample Estimation and Hypothesis Testing, *Handbook of Econometrics*, Vol. 4, North Holland, p. Chapter 36.

Phillips, P. 2012, Folklore Theorems, Implicit Maps, and Indirect Inference, *Econometrica* **80**(1), 425–454.

Ramberg, J. and Schmeiser, B. 1975, An Approximate Method for Generating Asymmetric Random Variables, *Communications of the ACM* **17**(2), 78–82.

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

Rothenberg, T. 1971, Identification in Parametric Models, *Econometrica* **39:3**, 577–591.

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.

Sargan, J. D. 1983, Identification and Lack of Identification, *Econometrica* **51:6**, 1605–1633.

Tauchen, G. 1998, The Objective Function of SImulation Estimators Near the Boundary of the Unstable Region of the Parameter Space, *Revuew of Economics and Statistcs* **80**, 389–398.

Tugnait, J. 1986, Identification of Non-Minimum Phase Linear Stochastic Systems, *Automatica* **22**, 457–464.

Tugnait, J. 1995, Parameter Estimation for Noncausal ARMA Models of Non-Gaussian Signals via Cumulant Matching, *IEEE Transactions on Signal Processing* **43**(4), 886–894.

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

θ_0	GMM estimator				Gaussian QML estimator				infeasible GMM estimator			
	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) distribution 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.

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) distribution 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 $m_{SMM,t}$, defined in (14), and the SMD estimator is based on the auxiliary parameter vector ψ_{SMD}, defined in (16c). The SMM and SMD estimators use the optimal weighting matrix based on the Newey-West HAC estimator.

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

errors/estimator	θ				α		
	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.

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

		skewness	kurtosis	QML	SMM	SMD
small	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)
	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)
2	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)
	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)
3	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	−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)
4	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)
	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)
big	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)
	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 $e_t \sim iid(0, \sigma^2)$ 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.

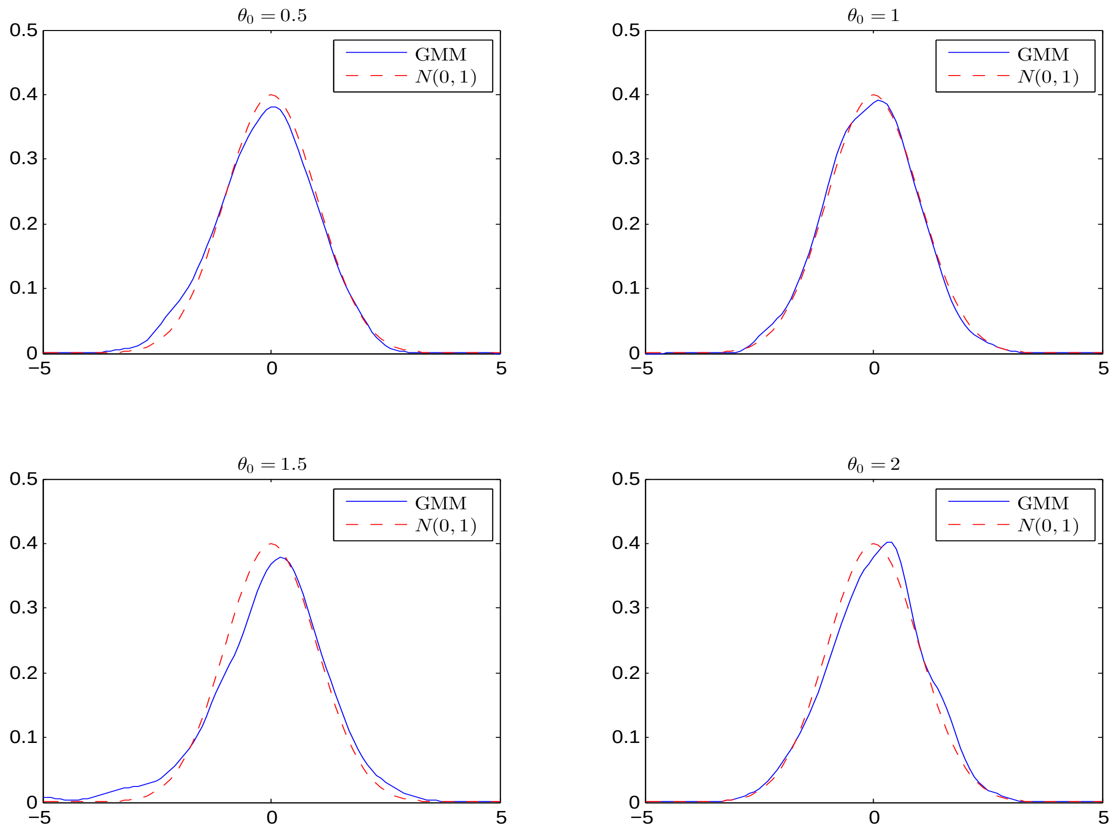

Figure 1: Density functions of the standardized GMM estimator (t-statistic) of θ based on data ($T = 3000$) generated from an MA(1) model $y_t = e_t + \theta e_{t-1}$ with $\theta = 0.5, 1, 1.5, 2$, and $e_t \sim iid(0, 1)$. The errors are drawn from a generalized lambda distribution with zero excess 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.

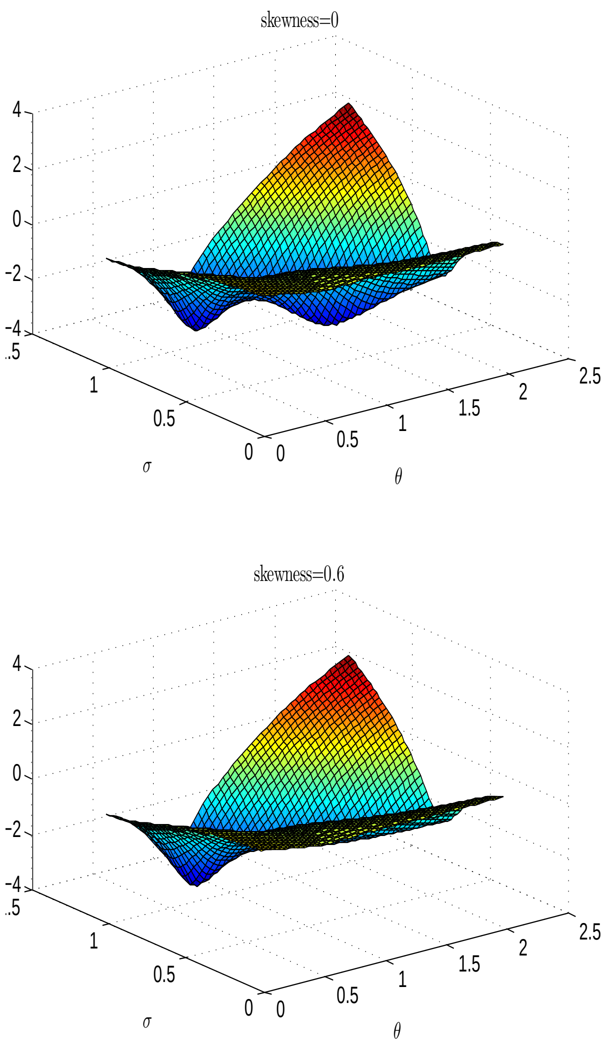

Figure 2: Logarithm of the objective function of simulation-based estimator of θ and σ based on data ($T = 1000$) generated from an MA(1) model $y_t = e_t + \theta e_{t-1}$ with $\theta = 0.7$ and $e_t \sim iid(0, 1)$. The errors are drawn from a generalized lambda distribution with zero excess kurtosis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