A Comparison of Minimum MSE and Maximum Power for the Nearly Integrated Non-Gaussian Model

Karim M. Abadir
André Lucas

A COMPARISON OF MINIMUM MSE AND MAXIMUM POWER FOR THE NEARLY INTEGRATED NON-GAUSSIAN MODEL

By Karim M. Abadir and André Lucas

University of York and Vrije Universiteit Amsterdam

We study the optimal choice of quasi-likelihoods for nearly integrated, possibly non-normal, autoregressive models. It turns out that the two most natural candidate criteria, minimum Mean Squared Error (MSE) and maximum power against the unit root null, give rise to different optimal quasi-likelihoods. In both cases, the functional specification of the optimal quasi-likelihood is the same: it is a combination of the true likelihood and the Gaussian quasi-likelihood. The optimal relative weights, however, depend on the criterion chosen and are markedly different. Throughout, we base our results on exact limiting distribution theory. We derive a new explicit expression for the joint density of the minimal sufficient functionals of Ornstein-Uhlenbeck processes, which also has applications in other fields, and we characterize its behaviour for extreme values of its arguments. Using these results, we derive the asymptotic power functions of statistics which converge weakly to combinations of these sufficient functionals. Finally, we evaluate numerically our computationally-efficient formulae.

1. Introduction. We study the optimal choice of quasi-likelihoods for inference in the autoregressive (AR) model

$$
(1) \qquad y_i = \phi y_{i-1} + \varepsilon_i
$$

where $\varepsilon_i \sim IID(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon^2 < \infty$. We assume that the (unknown) p.d.f. of ε_i exists and is given by $f(\varepsilon_i)$. We will be particularly interested in the case where ϕ is equal or close to unity, see Chan and Wei (1987) and Phillips (1987). We condition on y_0, and we do not consider explicitly higher order AR terms or additional short term dynamics. Phillips (1987) has shown how to adjust for these by imposing the appropriate mixing conditions on ε_i and correction factors such that our subsequent analysis is unaffected.

Much of the literature focuses on the case of Gaussian ε_i and the likelihood principle, or alternatively possibly non-Gaussian ε_i and the Gaussian quasi-likelihood principle. If innovations are truly Gaussian, Elliott, Rothenberg, and

---

Stock (1996) show that common tests for $H_0 : \phi = 1$ have a power that almost coincides with the Neyman-Pearson pointwise power envelope. Typical examples of such tests include the Studentized t statistic and the normalized autocorrelation. By contrast, Rothenberg and Stock (1997) show that under non-normality the same tests clearly display suboptimal power against a broad range of point alternatives, if the true likelihood is used to construct the tests. Therefore, there appears to be room for further improvement of the usual testing procedures when the true likelihood is non-Gaussian.

Our results complement earlier work by Cox and Llatas (1991). They consider the optimal choice of quasi-likelihoods from an asymptotic MSE perspective. Based on MSE, the optimal quasi-likelihood is a linear combination of the true likelihood and the Gaussian quasi-likelihood. The weight assigned to the Gaussian quasi-likelihood becomes negligible for distant alternatives. In the context of testing for $H_0 : \phi = 1$, however, it might be equally natural to consider power-related optimality criteria rather than MSE. We show that the form of the optimal quasi-likelihood in this case is analogous: it is a linear combination of the true and the Gaussian quasi-likelihood. Surprisingly, however, the weights differ markedly from those of Cox and Llatas. In particular, the weight of the Gaussian quasi-likelihood remains far from zero for a wide range of alternatives, thus impairing the optimality of the maximum likelihood estimator for these alternatives. Moreover, the discrepancy between the MSE-optimal weights and the power-optimal weights is a reminder that optimal estimation and optimal testing lead to different statistical procedures, which are not even asymptotically equivalent in our context.

The plan of the paper is as follows. In Section 2, we set out the basic power optimization problem for tests of $H_0 : \phi = 1$ in the framework of nearly integrated non-Gaussian time series. This leaves us with a one-dimensional maximization problem under one constraint. The first order conditions of this problem involve expectations of two functionals of an Ornstein-Uhlenbeck process. We exploit the notion that these functionals jointly form a minimal sufficient statistic for this process (which belongs to a curved exponential family) to derive analytically a new explicit expression for their joint density in Section 3. The formula for this density can also have applications in fields beyond the subject of our paper. The exact expression for the density is complemented with asymptotic expansions which characterize its behaviour for extreme (small or large) values of its arguments. Using the explicit form of the density, we are able to compute power functions for the Studentized t statistic and the normalized autocorrelation coefficient, up to any desired accuracy, and their asymptotic approximations too. The power functions are used to quantify the optimal choice of quasi-likelihoods in Section 4. We compare the optimal choices from a maximum power perspective (using both statistics) and a minimum MSE perspective. The true maximum-likelihood estimator is suboptimal in both cases. An important distinction emerges between optimality from a power and an MSE point of view. Some concluding remarks can be found in Section 5. In order not to interrupt the exposition, the proofs and some additional notation are gathered in an Appendix.

Throughout this paper we make use of the following notation. We use $\sim$ for denoting the distribution of a random variate, e.g., $\varepsilon_t \sim N(0,1)$, or for denoting asymptotic equivalence, e.g., $(1+x)^{-1} \sim x^{-1}$ for large $|x|$. The symbols $\xrightarrow{w}$ and $\xrightarrow{p}$ denote weak convergence and convergence in probability, respectively. The signum (sign) function is written as $\mathrm{sgn}(x)$. Expectations are denoted by $\mathbb{E}[\cdot]$ and probabilities by $\mathbb{P}[\cdot]$. Finally, $\phi(\cdot)$ and $\Phi(\cdot)$ are used to denote the standard normal p.d.f. and c.d.f., respectively.

2.Power optimization.   Let $-q(\cdot)$ denote the logarithm of a quasi-density used for estimating the autoregressive parameter ϕ_n. As explained in the introduction, we work in a local-to-unity framework. In particular, following Chan and Wei (1987) and Cox and Llatas (1991), we assume that $c = n \cdot (\phi - 1)$ is constant in the sample size n. Taking a locally quadratic approximation to the log-quasi-likelihood, see Rothenberg and Stock (1997), we obtain

$$(2) \qquad \log \mathcal{Q}(c, c_0; q) = -\sum_{i=1}^{n} q\left(\nabla y_i - \frac{c_0}{n} y_{i-1}\right)$$

$$= -\sum_{i=1}^{n} q\left(\varepsilon_i + \frac{c - c_0}{n} y_{i-1}\right)$$

$$= -\sum_{i=1}^{n} \left[q(\varepsilon_i) + \frac{c - c_0}{n} y_{i-1} q'(\varepsilon_i) + \right.$$

$$\left. \frac{(c - c_0)^2}{2n^2} y_{i-1}^2 q''(\varepsilon_i) \right] + r_n(c),$$

where $r_n(c)$ is the remainder term, ∇ is the backward difference operator such that $\nabla y_i = y_i - y_{i-1}$, and q' and q'' are the first and second derivatives of q, respectively.

We make the following assumption.

ASSUMPTION 1.   *(i) The innovations ε_i are i.i.d. with density $f(\cdot)$. (ii) The density $f(\cdot)$ is twice continuously differentiable and vanishes on the edge of its support. (iii) The log-quasi-density $q(\cdot)$ is twice continuously differentiable and $q''(\cdot)$ satisfies a linear Lipschitz condition. (iv) $(\varepsilon_i, q'(\varepsilon_i))$ has zero mean and finite covariance matrix.*

As mentioned in the introduction, the i.i.d. assumption can be relaxed by imposing the appropriate mixing conditions on ε_i, see Lucas (1995) and Elliott et al. (1996). The smoothness of the density in (ii) ensures that the maximum likelihood estimator and the information matrix are well defined. The assumption on the tails of $f(\cdot)$ makes it possible to replace $\int q''(\varepsilon) f(\varepsilon) d\varepsilon$ by $-\int q(\varepsilon) f'(\varepsilon) d\varepsilon$ and can be dispensed with at an increased complexity in the derivations. Con-

dition (iii) is a standard condition in this context of quasi-maximum likelihood estimators and nearly non-stationary time series, see, e.g., Lucas (1995).

Our main interest in this paper is in the optimal choice of the quasi-likelihood q. In particular, we will study the impact of different optimality criteria on the choice of q. Therefore, for simplicity we follow Rothenberg and Stock (1997) and abstain from estimating the variance σ_ε^2 of the innovations. Note that consistent estimators of σ_ε are asymptotically ancillary for ϕ, so that the choice of any consistent method of estimating σ_ε does not affect our subsequent derivations.

Using Assumption 1, the following theorem can be derived from Rothenberg and Stock (1997) for the limit of the quasi-likelihood ratio of testing $c = c_0$ against the fixed alternative $c = c_1$.

THEOREM 1 (ROTHENBERG & STOCK).   *Given Assumption 1,*

$$(3) \qquad n^{-1} \sum_{i=1}^{n} y_{i-1} q'(\varepsilon_i) \stackrel{w}{\to} \sigma_\varepsilon \sigma_q \cdot (\rho_q R + \xi\sqrt{(1-\rho_q^2)}S),$$

and

$$(4) \qquad n^{-2} \sum_{i=1}^{n} y_{i-1}^2 q''(\varepsilon_i) \stackrel{w}{\to} \sigma_\varepsilon^2 \delta_q S,$$

where $\sigma_\varepsilon^2 = \mathbb{E}[\varepsilon_i^2]$, $\sigma_q^2 = \mathbb{E}[q(\varepsilon_i)^2]$, $\delta_q = \mathbb{E}[q''(\varepsilon_i)]$, $\rho_q = \mathbb{E}[\varepsilon_i q'(\varepsilon_i)]/(\sigma_\varepsilon \sigma_q)$, $R = \int_0^1 U_c(\tau)dB_1(\tau)$, $S = \int_0^1 U_c(\tau)^2 d\tau$, B_1 is a standard Brownian motion, U_c is the Ornstein-Uhlenbeck process generated by the boundary condition $U_c(0) = 0$ and

$$(5) \qquad dU_c(\tau) = cU_c(\tau) d\tau + dB_1(\tau),$$

and ξ is a standard normal random variate independent of B_1.
Moreover, the likelihood ratio for testing $c = c_0$ against $c = c_1$ converges weakly as

$$QLR(c, c_0, c_1; q) \equiv \log \mathcal{Q}(c, c_0; q) - \log \mathcal{Q}(c, c_1; q)$$

$$(6) \qquad \stackrel{w}{\to} (c_0 - c_1)\sigma_q \sigma_\varepsilon \left(\rho_q R + \xi\sqrt{(1-\rho_q^2)}S - \frac{(c_0 + c_1 - 2c)\delta_q \sigma_\varepsilon S}{2\sigma_q} \right).$$

We are interested in the case $c_0 = 0$, i.e., the null of a unit root $H_0 : \phi = 1$. The upper bound for the power of any unit root test in this context, i.e., the power envelope, follows from the Neyman-Pearson lemma and is given by $\mathbb{P}[QLR(c_1, 0, c_1; f) < \kappa]$, where f is the unknown true density of the innovations ε_i, and $\mathbb{P}[QLR(0, 0, c_1; f) < \kappa]$ is equal to some prespecified significance level, see also Rothenberg and Stock (1997). Using the power envelope, we can evaluate the power performance of commonly employed unit root tests like the Studentized t statistic

$$(7) \qquad t_q = \frac{\hat{\phi}_q - 1}{\hat{\sigma}_q},$$

and the normalized autocorrelation coefficient,

$$(8) \qquad n \cdot (\hat{\phi}_q - 1),$$

where $\hat{\phi}_q$ denotes an estimator of ϕ based on the log-quasi-density $q(\cdot)$, and where $\hat{\sigma}_q^{-2}$ is an estimator of minus the Hessian (observed information). The usual choice for q is the Gaussian density, see Fuller (1976). Non-Gaussian quasi-densities have been used as well, see, e.g., Cox and Llatas (1991), Lucas (1995), and Rothenberg and Stock (1997).

It is useful to remark at this stage that the power envelope does not coincide with the power of the likelihood ratio (LR) test statistic. In particular, the LR test has the same asymptotic distribution under local alternatives as the squared Studentized t statistic of (7). As we show in Section 4, the Studentized t statistic does not attain the power envelope uniformly in c_1. Given the relation between the Studentized t statistic and the LR test, this also holds for the latter.

Rothenberg and Stock (1997) show that for the non-Gaussian case the usual tests in (7) and (8), though very close to the power envelope for some $c_1 \in (-\infty, 0]$, are considerably below the power envelope for other values of c_1 if the true likelihood is used, i.e., if $q(\cdot) = -\log f(\cdot)$. This sharply contrasts with the findings for the Gaussian case, see Elliott et al. (1996). Therefore, there appears to be room for further improvement of the usual testing procedures in (7) and (8) by allowing the quasi-likelihood used in there to differ from the true likelihood.

We formalize this as follows. Given Theorem 1, we consider statistics that can be written in the form

$$(9) \qquad T(R, S, \xi; \rho_q, \delta_q, c, \sigma_q, \sigma_\varepsilon).$$

For example, the Studentized t statistic in (7) can be written as

$$(10) \qquad t_q = \text{sgn}(\delta_q) \frac{\rho_q R}{\sqrt{S}} + \xi \sqrt{1 - \rho_q^2} + \frac{|\delta_q| \sigma_\varepsilon \sqrt{S}}{\sigma_q} c + o_p(1),$$

while the normalized autocorrelation equals

$$(11) \qquad n \cdot (\hat{\phi} - 1) = \frac{\sigma_q}{\sigma_\varepsilon \delta_q} \frac{\rho_q R + \xi \sqrt{(1 - \rho_q^2) S}}{S} + c + o_p(1);$$

see Cox and Llatas (1991), Lucas (1995), and Rothenberg and Stock (1997). For a given statistic T, we now want to maximize

$$(12) \qquad \mathbb{P}[T(R, S, \xi; \rho_q, \delta_q, c_1, \sigma_q, \sigma_\varepsilon) \in \mathcal{T}],$$

subject to

$$(13) \qquad \mathbb{P}[T(R, S, \xi; \rho_q, \delta_q, 0, \sigma_q, \sigma_\varepsilon) \in \mathcal{T}] = \alpha,$$

for some fixed level of the test α, where $\mathcal{T}$ is the critical region of the test. Usually, $\mathcal{T}$ is of the form $(-\infty, \kappa]$ for some κ, see for example the two statistics in (10) and (11). The optimization of (12)–(13) is carried out with respect to the log-quasi-density $q(\cdot)$.

Note that $q(\cdot)$ does not enter the equations (12) and (13) directly. Only q' enters through ρ_q and σ_q, while q'' enters through δ_q. This allows us to simplify the optimization problem by optimizing with respect to q' rather than q under the additional constraint that $\delta_q = 1$. The latter constraint is a simple normalization constraint which reduces the complexity of the problem without affecting the resulting optimal choice of the estimator. This is easily seen by noting that the QML estimator is defined by the first-order condition

$$\sum_{i=1}^{n} q'(y_i - \hat{\phi}y_{i-1}) = 0.$$

The estimator, therefore, does not change if q' is multiplied by a constant δ_q^{-1}. The Lagrangian of the optimization problem now becomes

$$(14) \qquad \max_{q'} \mathbb{P}^*(\rho_q, c, \sigma_q, \sigma_\varepsilon, \lambda_0) + \lambda_0 \alpha - \lambda_1 \mathbb{E}[q'(\varepsilon_i)] + \lambda_2(1 - \delta_q),$$

with λ_i the Lagrange multipliers of the appropriate constraints, and

$$\mathbb{P}^*(\rho_q, c, \sigma_q, \sigma_\varepsilon, \lambda_0) = \mathbb{P}[T(R, S, \xi; \rho_q, \delta_q, c, \sigma_q, \sigma_\varepsilon) \in \mathcal{T}] -$$

$$\lambda_0 \mathbb{P}[T(R, S, \xi; \rho_q, \delta_q, 0, \sigma_q, \sigma_\varepsilon) \in \mathcal{T}].$$

The first-order conditions of (14) are given by the Euler-Lagrange equations

$$(15) \qquad 0 = \frac{\sigma_q \varepsilon f - \rho_q \sigma_\varepsilon q' f}{\sigma_\varepsilon \sigma_q^2} \cdot \frac{\partial}{\partial \rho_q} \mathbb{P}^*(\rho_q, c, \sigma_q, \sigma_\varepsilon, \lambda_0) +$$

$$\frac{q' f}{\sigma_q} \cdot \frac{\partial}{\partial \sigma_q} \mathbb{P}^*(\rho_q, c, \sigma_q, \sigma_\varepsilon, \lambda_0) - \lambda_1 f + \lambda_2 f',$$

$$(16) \qquad \alpha = \mathbb{P}(T(R, S, \xi; \rho_q, 1, 0, \sigma_q, \sigma_\varepsilon) \in \mathcal{T}),$$

$$(17) \qquad 0 = \mathbb{E}[q'(\varepsilon_i)],$$

$$(18) \qquad 1 = \delta_q.$$

Integrating (16) and using (17) and $\mathbb{E}[\varepsilon_i] = 0$, we obtain $\lambda_1 = 0$. It is then easy to see from (16) that the optimal q' must be of the form

$$(19) \qquad q'(\varepsilon) = m_1 \cdot \varepsilon - m_2 \cdot \mathcal{I}^{-1} f'(\varepsilon)/f(\varepsilon),$$

with information $\mathcal{I} \equiv \mathbb{E}[(f'(\varepsilon)/f(\varepsilon))^2]$, such that the optimal quasi-score $q'(\varepsilon)$ is a linear combination of the Gaussian quasi-score and the true likelihood score with weights m_1 and m_2, respectively. As mentioned earlier, a comparable result was obtained by Cox and Llatas (1991) using a minimum MSE criterion rather than maximum power.

Using (18), (19), and the information matrix equality, we obtain

$$(20) \qquad m_1 - m_2 \mathcal{I}^{-1} \mathbb{E}[d^2 \log f(\varepsilon)/d\varepsilon^2] = m_1 + m_2 = 1.$$

Note that (17) is trivially satisfied by (19). So instead of maximizing over a function space (with respect to q'), we can do the much simpler maximization with respect to the scalar m_1. Using (19), we also obtain:

$$(21) \qquad \tilde{\sigma}_q^2 \equiv \sigma_q^2/\sigma_\varepsilon^2 = m_1^2 + 2m_1 m_2 \tilde{\mathcal{I}}^{-1} + m_2^2 \tilde{\mathcal{I}}^{-1} = m_1^2(1 - \tilde{\mathcal{I}}^{-1}) + \tilde{\mathcal{I}}^{-1},$$

and

$$(22) \qquad \rho_q = [m_1(\tilde{\mathcal{I}} - 1) + 1]/\sqrt{\tilde{\mathcal{I}} \cdot (m_1^2(\tilde{\mathcal{I}} - 1) + 1)},$$

where $\tilde{\mathcal{I}} = \sigma_\varepsilon^2 \mathcal{I} \geq 1$ with equality only for the normal density, see Rustagi (1976).

For the remainder of this paper, we focus on the Studentized t statistic and the normalized autocorrelation as test statistics for $H_0 : \phi = 1$. For both of these, the critical region $\mathcal{T}$ is of the form $(-\infty, \kappa]$. Because ξ is symmetrically distributed, we can assume without loss of generality that these statistics are monotonically increasing in ξ, except at $\rho_q = 1$. Also note that the Studentized t statistic and normalized autocorrelation coefficient only depend on $\tilde{\sigma}_q$ instead of σ_q and σ_ε separately, see (10) and (11). It is therefore possible to define $T_*^{-1}(R, S, \kappa; \rho_q, c, \tilde{\sigma}_q)$ as the upper endpoint of the region $T^{-1}(R, S, \mathcal{T}; \rho_q, 1, c, \sigma_q, \sigma_\varepsilon)$, where $T^{-1}(\cdot)$ is the inverse of the statistic with respect to ξ. If we further define $h_c(r, s)$ as the density of (R, S) for a given value of c, the first order conditions of the maximization problem simplify to

$$(23) \qquad \left[\frac{(1 - m_1)}{m_1^2(\tilde{\mathcal{I}} - 1) + 1} \frac{\partial}{\partial \rho_q} + m_1 \frac{\partial}{\partial \tilde{\sigma}_q} \right] \cdot \int \int \left[\Phi(T_*^{-1}(r, s, \kappa; \rho_q, c_1, \tilde{\sigma}_q)) - \right.$$

$$\left. \lambda_0 \Phi(T_*^{-1}(r, s, \kappa; \rho_q, 0, \tilde{\sigma}_q)) \right] \cdot h_c(r, s) dr ds = 0,$$

$$(24) \qquad \int \int \Phi(T_*^{-1}(r, s, \kappa; \rho_q, 0, \tilde{\sigma}_q)) h_0(r, s) dr ds = \alpha,$$

where

$$(25) \qquad T_*^{-1}(r, s, \kappa; \rho_q, c, \tilde{\sigma}_q) = \frac{\kappa - \rho_q r/\sqrt{s} - c\sqrt{s}/\tilde{\sigma}_q}{\sqrt{1 - \rho_q^2}}$$

for the Studentized t statistic, and

$$(26) \qquad T_*^{-1}(r, s, \kappa; \rho_q, c, \tilde{\sigma}_q) = \frac{-\rho_q r/\sqrt{s} + (\kappa - c)\sqrt{s}/\tilde{\sigma}_q}{\sqrt{1 - \rho_q^2}}$$

for the normalized autocorrelation. It is clear that in order to solve (23)–(24), we need a way to accurately evaluate the integral. More specifically, we need an expression for the density $h_c(r, s)$. This is the subject of the next section.

3.Exact distributions and their large-c approximations. It is useful to obtain an explicit expression for the joint density $h_c(r,s)$. Not only can it help us obtain the desired optimal weights m_1 for our problem, but also Perron (1989,1991) has shown how numerical evaluation of the marginal distribution for R/S can be useful to approximate accurately the power function of the normalized least-squares estimator of ϕ in the standard Gaussian case. The power function for the Studentized t statistic would be based on $R/\sqrt{S}$, see (10). Its computation would have involved triple integration (two of them in the complex space) in the standard Gaussian case, and would have required two more integrals (five in all) in our case. For this reason, neither case has been derived explicitly in the literature.

Let

$$r_+ \equiv r + cs + \frac{1}{2} \in \mathcal{R}_+.$$

We then have the following result.

THEOREM 2. *The joint density function of* $\left(\int_0^1 U_c(\tau)\, dB(\tau), \int_0^1 U_c(\tau)^2\, d\tau \right)$ *is*

$$(27) \qquad h_c(r,s) = \exp\left[cr + \frac{c^2}{2}s \right] h_0(r+cs,s)$$

$$= \frac{\exp\left[c\left(r_+ - \frac{1}{2}\right) - \frac{r_+^2}{2s} - \frac{c^2 s}{2} \right]}{s^{\frac{5}{4}}\sqrt{\pi r_+}} \sum_{j=0}^{\infty} \binom{j - \frac{1}{2}}{j} \cdot$$

$$\exp\left[-\left(2j + \frac{1}{2}\right)\frac{r_+}{s} - \left(2j^2 + j + \frac{1}{8}\right)\frac{1}{s} \right]$$

$$\sum_{\ell=0}^{j} \binom{j}{\ell} \frac{(-2r_+/\sqrt{s})^{\ell}}{\Gamma\left(\ell + \frac{1}{2}\right)} D^+_{\ell+\frac{3}{2}}\left(\frac{r_+ + 2j + \frac{1}{2}}{\sqrt{s}} \right),$$

where $D^+_\nu(z) \equiv \exp(z^2/4) D_\nu(z)$, *and* $D_\nu(z)$ *is a parabolic cylinder function, see Erdélyi (1953) for details.*

This expansion is very efficient. The sum in j typically converges extremely rapidly (in one or two terms) because of the exponential factor in $-j^2$, while the sum in ℓ is bounded by j. For more details on computational issues, see Abadir (1992,1995). Efficient power series for D^+ have been derived in Abadir (1993), and are summarized in Abadir (1999).

We are also interested in characterizing how this density behaves as (r,s) tend to extreme values, especially since we wish to assess the power functions as c varies, both exactly and approximately. In particular we will be using the fact in Phillips (1987) that, as $c \to -\infty$,

$$(28) \qquad \sqrt{-2c}R = \sqrt{-2c}\int_0^1 U_c(\tau)\, dB(\tau) \overset{w}{\to} N(0,1)$$

(29) $$-2cS = -2c \int_0^1 U_c(\tau)^2 \, d\tau \xrightarrow{p} 1,$$

which will indicate how the density behaves in the case of distant stable alternatives. Note that (29) implies that S becomes ancillary for ϕ as $c \to -\infty$, in which case the statistical curvature (and the inferential complications it causes) of this exponential-family process will vanish. We have the following corollary.

COROLLARY 1.   *The asymptotic expansion of $h_c(r,s)$ is given by:*

$$\frac{\left(\frac12 + r_+\right)^{\frac32}}{\pi s^2 \sqrt{r_+}} \exp\left[c\left(r_+ - \frac12\right) - \frac{c^2 s}{2} - \frac{4r_+^2 + 4r_+ + 1}{8s} \right],$$

when either $r_+/s, 1/s \to \infty$;

$$\frac{1}{\pi} \sqrt{\frac{3}{sr_+}} \exp\left[-\frac{c}{2} + (c-1)r_+ - \frac{6+c^2}{2}s \right],$$

when both $r_+/s, 1/s \to 0$;

$$\frac{1}{\pi s^{\frac54} \sqrt{r_+}} \exp\left[-\frac{c}{2} - \frac{c^2 s}{2} \right] \sum_{j=0}^{\infty} \binom{j - \frac12}{j} D_{\frac32}^+ \left(\frac{2j + \frac12}{\sqrt{s}} \right) \exp\left[-\left(2j^2 + j + \frac18\right) \frac1s \right],$$

when $r_+ \to 0$.

We have included all cases for the sake of completeness, but the part that will be most relevant for our subsequent manipulations as $c \to -\infty$ is the first expression of the corollary, as is clear from (28) and (29). Asymptotic expansions, such as the ones in this corollary and the next, are usable for large but not necessarily infinite values of the parameters.

We now need to obtain the asymptotic power functions of the Studentized t statistic and the normalized autocorrelation as

$$\Pi_t\left(\kappa, c, \rho_q, \tilde\sigma_q\right) \equiv \iint\limits_{\mathcal{R}_+^2} \Phi\left(\frac{\kappa - \rho_q r/\sqrt{s} - c\sqrt{s}/\tilde\sigma_q}{\sqrt{1 - \rho_q^2}} \right) h_c(r,s) \, ds \, dr_+,$$

and

$$\Pi_\phi\left(\kappa, c, \rho_q, \tilde\sigma_q\right) \equiv \iint\limits_{\mathcal{R}_+^2} \Phi\left(\frac{(\kappa - c)\sqrt{s}/\tilde\sigma_q - \rho_q r/\sqrt{s}}{\sqrt{1 - \rho_q^2}} \right) h_c(r,s) \, ds \, dr_+,$$

respectively. Since we now have an efficient explicit formula for $h_c(r,s)$, this can be done to any required numerical precision by double numerical integration. We do so in our next section on numerical results. Here, we consider one further manipulation, namely how the two $\Pi\left(\kappa, c, \rho_q, \tilde\sigma_q\right)$ behave as $c \to -\infty$.

COROLLARY 2.   *Let*

$$\gamma \equiv \left(\left(1 - \rho_q^2\right) \tilde{\sigma}_q^2 + \left(1 - \rho_q \tilde{\sigma}_q\right)^2 \right)^{\frac{1}{4}} = \left(1 - \tilde{\mathcal{I}}^{-1}\right)^{\frac{1}{4}} \sqrt{|m_1 - 1|},$$

$$\gamma_c \equiv \left(\left(1 - \rho_q^2\right) \tilde{\sigma}_q^2 + \left(1 - \rho_q \tilde{\sigma}_q - \frac{\kappa}{c}\right)^2 \right)^{\frac{1}{4}} \sim \gamma,$$

and $\mathrm{Re}\{\cdot\}$ *denote the real part of an expression. As* $c \to -\infty$, *the complement of the power function of the* t *statistic,* $1 - \Pi_t\left(\kappa, c, \rho_q, \tilde{\sigma}_q\right)$, *is*

$$\mathrm{Re} \left\{ \frac{\exp\left[|c|\right]}{|\rho_q| \sqrt{2\rho_q^2 + 2|\rho_q|\left(\rho_q - \tilde{\sigma}_q\right) - 1}} \Phi\left(-\left(\mathrm{sgn}\left(\rho_q\right) + 1\right) \sqrt{\frac{|\rho_q c|}{\tilde{\sigma}_q}} - \kappa \right) \right.$$

$$+ \frac{\sqrt{2}\left(1 - \rho_q^2\right)\gamma}{\sqrt{\tilde{\sigma}_q - \rho_q + \left(1 - 2\rho_q^2\right)\gamma^2}} \Phi\left(\frac{\left(\tilde{\sigma}_q - \gamma^2\right)\rho_q - 1}{\gamma\sqrt{2\left(1 - \rho_q^2\right)\tilde{\sigma}_q}} \sqrt{|c|} - \frac{\kappa}{\sqrt{1 - \rho_q^2}} \right)$$

$$\left. \exp\left[\left(2\left(1 - \rho_q^2\right)\tilde{\sigma}_q + \left(\rho_q^2 - 2\right)\gamma^2 + \frac{\left(1 - \rho_q \tilde{\sigma}_q\right)^2}{\gamma^2} \right) \frac{|c|}{4\left(1 - \rho_q^2\right)\tilde{\sigma}_q} \right] \right\},$$

and that of the normalized autocorrelation, $1 - \Pi_\phi\left(\kappa, c, \rho_q, \tilde{\sigma}_q\right)$, *is*

$$\mathrm{Re} \left\{ \frac{\exp\left[|c|\right] \Phi\left(\left(\mathrm{sgn}\left(-\rho_q\right) + \mathrm{sgn}\left(\frac{\kappa}{c} - 1\right)\right) \sqrt{\frac{|(\kappa - c)\rho_q|}{\tilde{\sigma}_q}} \right)}{|\rho_q| \sqrt{2\rho_q^2 + 2\left|\frac{\rho_q c}{\kappa - c}\right| \left(\left(1 - \frac{\kappa}{c}\right)\rho_q - \tilde{\sigma}_q\right) - 1}} \right.$$

$$+ \frac{\sqrt{2}\left(1 - \rho_q^2\right)\gamma_c}{\sqrt{\tilde{\sigma}_q + \left(\frac{\kappa}{c} - 1\right)\rho_q + \left(1 - 2\rho_q^2\right)\gamma_c^2}} \Phi\left(\frac{\frac{\kappa}{c} + \left(\tilde{\sigma}_q - \gamma_c^2\right)\rho_q - 1}{\gamma_c\sqrt{2\left(1 - \rho_q^2\right)\tilde{\sigma}_q}} \sqrt{|c|} \right)$$

$$\left. \exp\left[\left(2\left(1 - \rho_q^2\right)\tilde{\sigma}_q + \left(\rho_q^2 - 2\right)\gamma_c^2 + \frac{\left(1 - \rho_q \tilde{\sigma}_q - \frac{\kappa}{c}\right)^2}{\gamma_c^2} \right) \frac{|c|}{4\left(1 - \rho_q^2\right)\tilde{\sigma}_q} \right] \right\}.$$

If required, one may approximate these expressions further by using

$$\Phi\left(z\right) \sim 1_{z \in \mathcal{R}_+} - \frac{\phi\left(z\right)}{z}, \qquad \text{as } z \to \pm\infty,$$

where $1_{z \in \mathcal{R}_+} = 1$ for $z \in \mathcal{R}_+$, and zero otherwise. Taking the real part of the sum of the two expressions involving $\Phi\left(\cdot\right)$ switches from one of the two components to the other one, depending on some parameter combinations. This discontinuity in the parameter space is typical of asymptotic expansions, and is known in Complex Analysis as Stokes' phenomenon. Note that $\rho_q \tilde{\sigma}_q$ is a weighted average of the estimator's relative precision $\tilde{\mathcal{I}}^{-1}$ and its value of 1 under a pseudo-Gaussian likelihood, namely

$$\rho_q \tilde{\sigma}_q = m_1 + \left(1 - m_1\right)\tilde{\mathcal{I}}^{-1},$$

while γ is a measure of deviation of $\tilde{\mathcal{I}}$ from 1, and of m_1 from 1.

The corollary points out that the asymptotic expansion of $1 - \Pi_t(\kappa, c, \rho_q, \tilde{\sigma}_q)$ is a shifted and rescaled Normal when viewed as a function of κ only, which is not generally the case for $1 - \Pi_\phi(\kappa, c, \rho_q, \tilde{\sigma}_q)$. However, as $c \to -\infty$, the influence of κ diminishes relative to that of c, and both power functions share the same expansion to first order. This is not surprising since S is asymptotically ancillary for ϕ as $c \to -\infty$ (see (29)) and the two statistics differ stochastically only by a factor of $\sqrt{S}$, see (10) and (11).

4. Results.   Given the exact and approximate distributional results of the previous section, we can now proceed by computing numerically the optimal quasi-likelihood. The integrals in (23) and (24) are evaluated using Gaussian quadrature. In order to mitigate potential numerical errors, we use the following change of variables:

$$(30) \quad \begin{cases} \tilde{r} = 1 - \exp(-\sqrt[4]{r + cs + \frac{1}{2}}) \\ \tilde{s} = 1 - \exp(-s) \end{cases} \Leftrightarrow \begin{cases} r = [\log(1 - \tilde{r})]^4 - \frac{1}{2} + c\log(1 - \tilde{s}) \\ s = -\log(1 - \tilde{s}) \end{cases}$$

giving rise to the Jacobian $4[-\log(1 - \tilde{r})]^3 (1 - \tilde{s})^{-1} (1 - \tilde{r})^{-1}$. This transformation accomplishes two things. First, Corollary 1 shows that the density $h_c(r, s)$ is $O(1/\sqrt{r + cs + 0.5})$ for $r \to -(cs + 0.5)$. This might be a source for numerical instabilities when integrating numerically near $r = -(cs + 0.5)$. Taking the 4th root $(\sqrt[4]{r + cs + 0.5})$ ensures that the density in the transformed variables $(\tilde{r}, \tilde{s})$ goes to zero as $r \to -(cs + 0.5)$. Secondly, the transformation maps the integration quadrant $(\sqrt[4]{r + cs + 0.5}, s) \in \mathcal{R}_+ \times \mathcal{R}_+$ into the unit square, thus preventing potential errors due to a cut-off of the tails over an unbounded region of integration.

Figures 1 and 2 plot several objective functions in m_1-space, in order to illustrate the power optimization exercise. The top two graphs give the power of the test statistics for different values of c and $\tilde{\mathcal{I}}$. In the lower two graphs, power is rescaled using the inverse normal cdf. This seems a sensible approach given the asymptotic expansions in Section 3. First note in the upper panels that the power function becomes increasingly flat for increasing values of $|c|$ and $\tilde{\mathcal{I}}$. Computing the optimal value of m_1 from a power perspective is, therefore, only possible if we have analytic and highly accurate expressions for the power function as derived in the previous section. More conventional simulation-based approaches would be hampered by simulation errors, given the high precision required to deal with our almost-flat objective function. A second thing to note in Figures 1 and 2 is the adequacy of the large-c expansion derived in Corollary 2. For increasing values of $|c|$, the match between the power based on numerical integration and on the asymptotic $(c \to -\infty)$ approximation becomes better. This is especially clear if we use a different scale on the vertical axis, as is done in the lower graphs. Also, the location of the optimal values of m_1 from both formulae (exact and approximate) are very close for large negative values of c. As discussed at the end of the previous section, by definition, asymptotic expansions do not provide
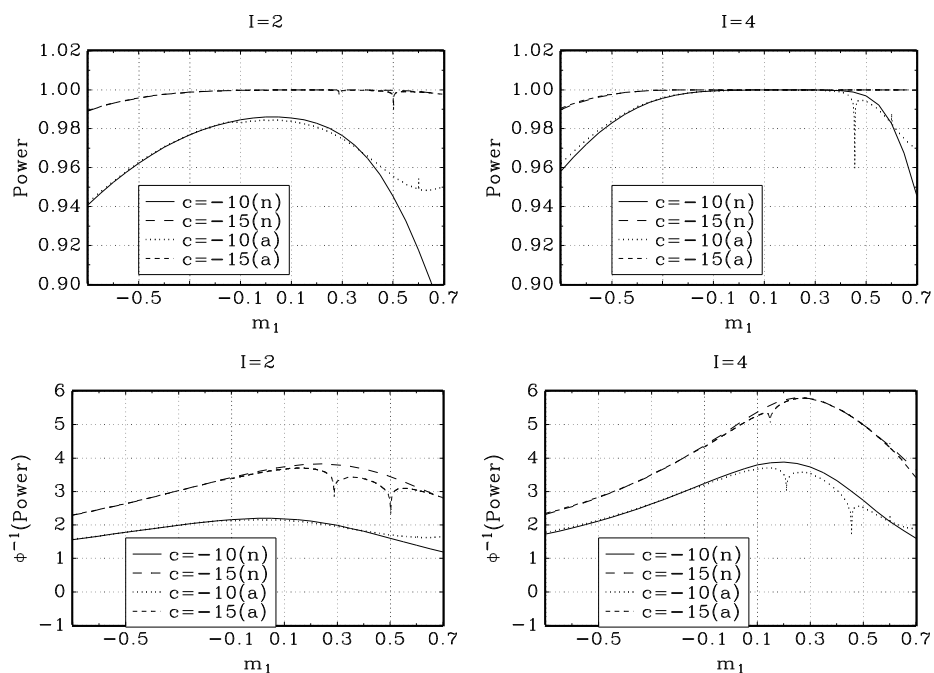
FIG. 1.   *Objective function for Studentized t statistic*
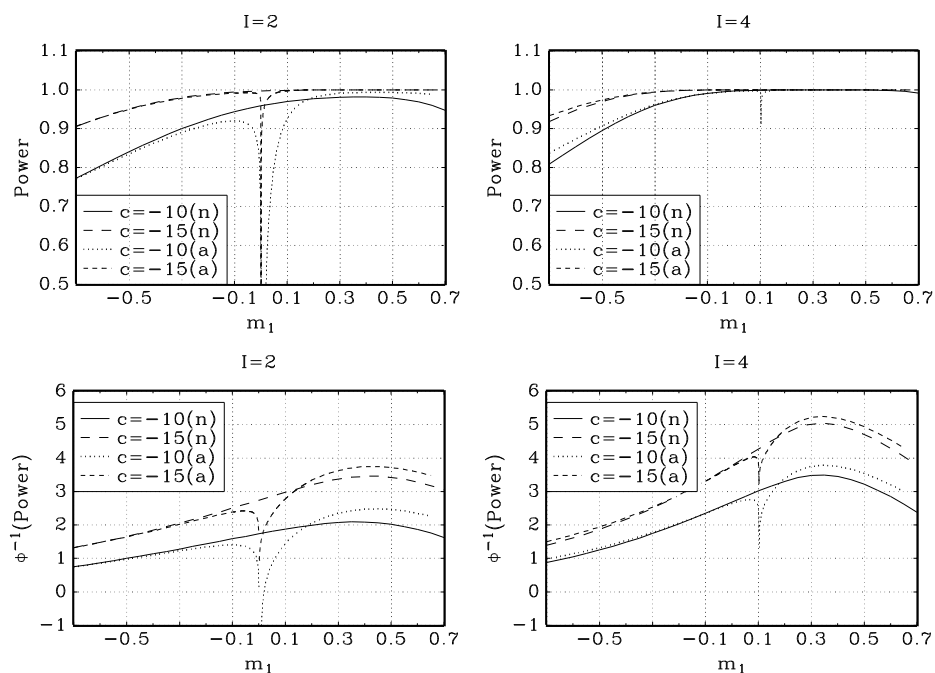
This figure contains the power of the Studentized t statistic for testing the unit root null at a 5% significance level. Power is displayed as a function of the Gaussian weight m_1 in the quasi score q'. The weight of the true likelihood score in q' is $1 - m_1$. The parameter c gives the distance between the stable alternative and the null of the unit root, while $I = \tilde{\mathcal{I}} \geq 1$ is an information measure of non-normality ($I = 1$ for the Gaussian case). The top two graphs have power on the vertical axis, while the lower two graphs have the inverse standard normal cdf of power on the vertical axis. Finally, the curves labeled (n) are obtained by numerical integration of the formulae in Theorem 2, while the curves labeled (a) are based on the asymptotic expansion in Corollary 2.

uniformly good approximations, and this is reflected in our graphs for some values of m_1. Note, however, that for large $|c|$ the expansion of Corollary 2 is still adequate in the region which is of interest to us in this context, i.e., near the optimum of the power function.

The results on the optimal values of m_1 for several values of $\tilde{\mathcal{I}}$ and c are displayed in Figure 3 for the Studentized t statistic and the normalized auto-correlation. For comparison, the optimal values of m_1 from a minimum MSE perspective are also provided using the approximations given in Cox and Llatas (1991). First note that the optimal m_1 values from a power perspective take on both positive and negative values (see the first two panels). By contrast, the

FIG. 2.   *Objective function for normalized autocorrelation*

This figure contains the power of the normalized autocorrelation statistic for testing the unit root null at a 5% significance level. Power is displayed as a function of the Gaussian weight m_1 in the quasi score q'. For a further description, see the explanatory note to Figure 1.

weights based on minimum MSE are uniformly negative. The point m_1 where the optimal weight crosses the horizontal axis denotes the point where the maximum likelihood estimator is optimal from a power perspective. A second point to note is that the weights m_1 in the Studentized t statistic for small values of $|c|$ are smaller than those for the normalized autocorrelation. Moreover, the magnitude of the weights of Cox and Llatas is even smaller. The third striking feature is that the optimal weights we compute lie persistently above the horizontal axis for distant alternatives. To corroborate this result, we use the asymptotic expansion of Corollary 2 to compute the optimal value of m_1 for extremely distant alternatives. Using the symbolic computation package Maple to obtain highly accurate (100 digits) approximations of the functions involved, we obtain the optimal values m_1 pictured in the lower-right graph in Figure 3. The weights clearly decrease as one would expect, albeit at a very slow rate. Also note from Figure 1 that the power function is very close to unity for these distant alternatives and that the objective function is very flat. Precise computations of

Fig. 3. *Optimal weights of Gaussian quasi score*

Autocorrelation

Studentized t

Cox–Llatas m_1
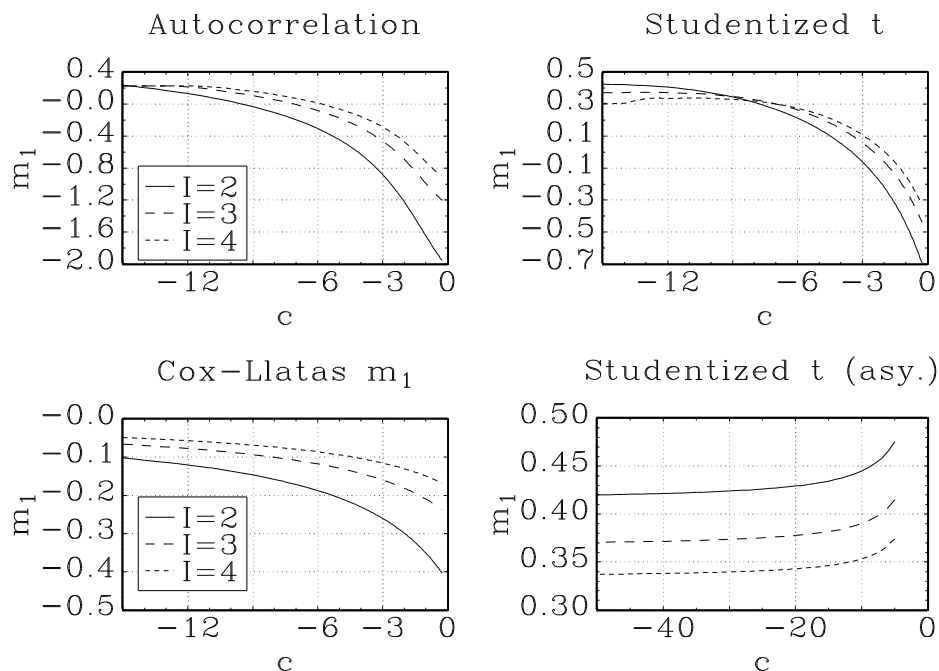
Studentized t (asy.)

This figure contains the optimal weights m_1 of the Gaussian quasi score in q'. The upper-left and upper-right panels give the results for maximum power using the normalized autocorrelation and the Studentized t statistic, respectively. The lower-left panel gives the optimal weights from a minimum MSE perspective, see Cox and Llatas (1991). Finally, the lower-right graph gives the optimal m_1 for the Studentized t statistic and a maximum power objective for distant alternatives using the asymptotic expansion of Corollary 2. The variable $I \equiv \tilde{\mathcal{I}} \geq 1$ is an information measure of non-normality ($I = 1$ for the Gaussian case).

the optimal m_1 for $c < -15$ are therefore only possible given the exact explicit expressions of Section 3.

We now turn to a discussion on the power behavior of the alternative testing procedures. The results are given in Figures 4 and 5. First we discuss the Studentized t statistic. For local alternatives ($c > -4$) all tests procedures have power close to the power envelope, see the top graphs. The percentage power loss with respect to the envelope has its maximum of around 13 to 14 percent for these local alternatives as well. Clearly, the optimal weights from the two top panels in our previous Figure 3 result in the smallest power loss with respect to the envelope. Two interesting features remain. First, even for distant alternatives one can gain power by choosing a quasi-likelihood different from the true likelihood. This is seen by the Abadir-Lucas curve in the top panels lying closer to the power

FIG. 4.   *Optimality results for the Studentized t statistic*
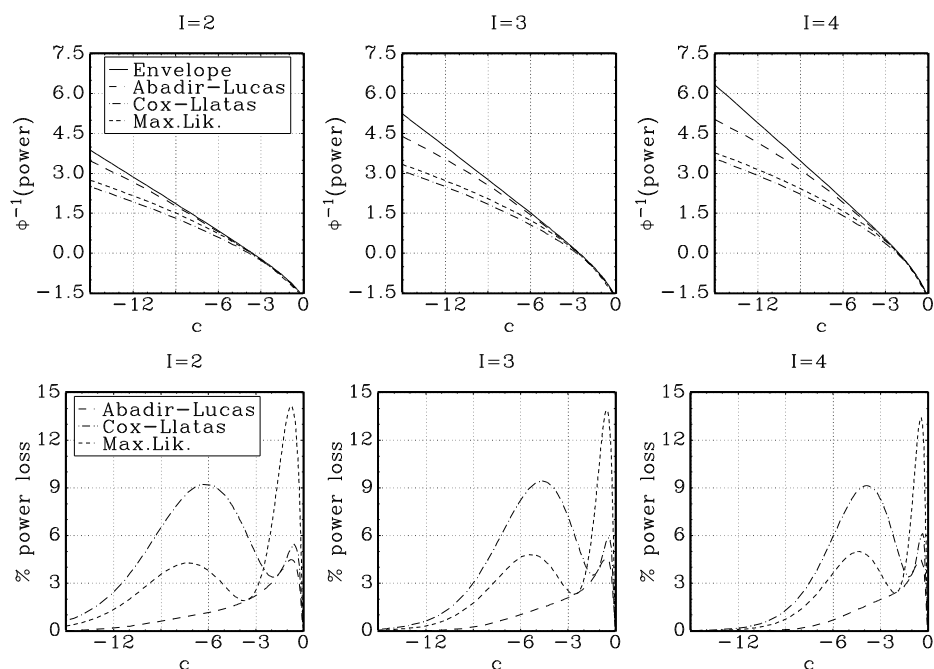
This figure contains the inverse normal c.d.f. of the power of the Studentized t statistic using the optimal weights m_1 from Figure 3. The Abadir-Lucas weight is that from the upper graphs in Figure 3. The Cox-Llatas weight m_1 is also given in Figure 3. The maximum likelihood estimator is given by $m_1 \equiv 0$. The lower graphs in the figure present the percentage power loss with respect to the power envelope. Again, $I = \tilde{\mathcal{I}}$ measures the non-normality, see the note to Figure 3.

envelope than the curves for the maximum likelihood estimator. Second, for values of c close to zero, the Studentized t statistic based on the weights of Cox and Llatas (minimum MSE) performs better than that based on the maximum likelihood estimator. For distant alternatives, however, the maximum likelihood approach dominates the Cox-Llatas one. Also note that both the Cox-Llatas and the maximum likelihood approach reveal a nonmonotonic percentage power loss for the Studentized t statistic.

For the normalized autocorrelation, the results are qualitatively similar. Again, the Abadir-Lucas curve in the upper panel lies closer to the power envelope than the maximum likelihood curve. The percentage power loss in the lower panels is again nonmonotonic. The second peak, however, is hardly visible due to the magnitude of the power loss for alternatives c close to zero. For the maximum likelihood, these power losses can amount from 35 up to 45 percent

FIG. 5. *Optimality results for the normalized autocorrelation*
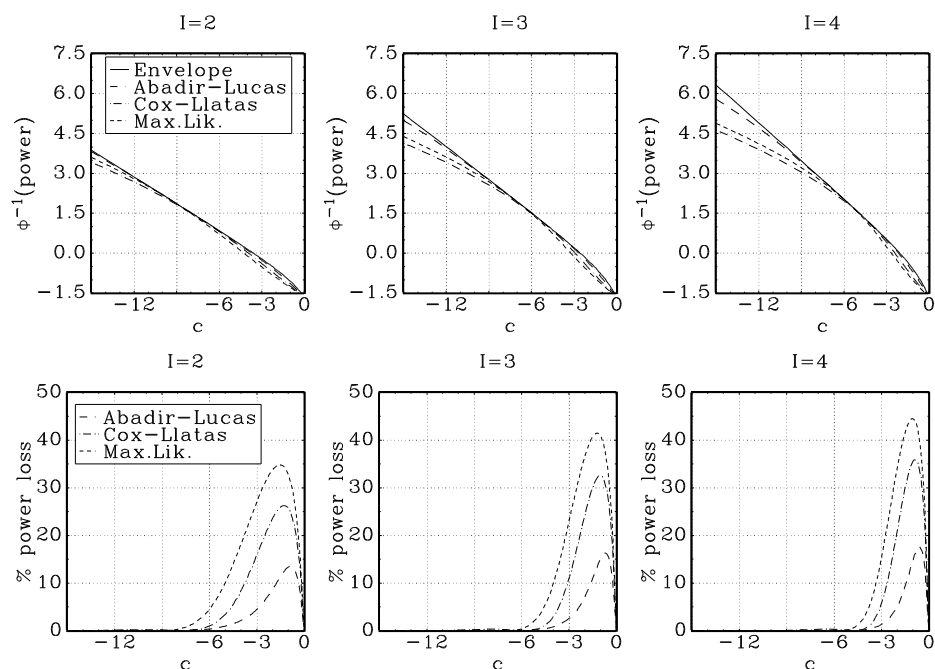
This figure contains similar information as Figure 4, but now for the normalized autocorrelation instead of the Studentized t statistic.

for c between -1 and 0. Also note that as expected the maximum likelihood approach dominates the approach of Cox-Llatas for $|c|$ sufficiently large, while the reverse holds for c sufficiently close to 0.

5. Concluding remarks.   In this paper we have quantified the difference between optimal estimators from a minimum MSE and a maximum power perspective for nearly non-stationary, non-Gaussian AR models. We have based our results on new explicit expressions for the joint density of the sufficient functionals of Ornstein-Uhlenbeck processes. This has allowed us to obtain highly accurate numerical results that could not have been obtained using traditional simulation techniques. The enhanced accuracy has allowed us to optimize different criterion functions based on MSE or on power against the unit root null over the space of possible quasi-likelihoods. In all cases considered, the optimal quasi-likelihood-score was a linear combination of the Gaussian quasi-score and the true likelihood-score. The weights, however, depended very much on the criterion chosen. In particular, minimum MSE and maximum power gave rise to

markedly different weights, illustrating how one should be additionally careful in this time series setting when choosing between different reasonable objectives for the design of optimal statistical procedures.

Appendix

Throughout, the change of a variable of integration which maps $u \mapsto v \equiv g(u)$ will be written in the inverse-mapping form $u \leftrightarrow g^{-1}(v)$, whereby u is replaced by $g^{-1}(v)$ in the integrand.

Proof of Theorem 2. For $c = 0$, the joint density function $h_0(r, s)$ has been derived in Abadir (1992, 1995) by inverting the joint characterisitic function of

$$\left(\int_0^1 B(\tau) \, dB(\tau), \int_0^1 B(\tau)^2 \, d\tau \right),$$

which was in White (1958). For general c, we can use the same procedure to invert the characteristic function given in Phillips (1987) and in Perron (1989) for the relevant functionals. It is shorter, however, to use exponential family and sufficiency arguments [e.g. see Lehmann (1986, Ch.2)] for this purpose. In the asymptotic context we are dealing with, Girsanov's (1960) theorem [see also Cameron and Martin (1944, 1945)] implies that the diffusion $U_c(\tau)$ is a Brownian motion with respect to another measure. The Radon-Nikodym derivative of the probability measure for U_c with respect to that for B is

$$\exp\left[c \int_0^1 U_c(\tau) \, dB(\tau) - \frac{c^2}{2} \int_0^1 U_c(\tau)^2 \, d\tau \right],$$

and we have that

$$\left(\int_0^1 U_c(\tau) \, dB(\tau), \int_0^1 U_c(\tau)^2 \, d\tau \right) = \left(\int_0^1 U_c(\tau) \, dU_c(\tau) - c \int_0^1 U_c(\tau)^2 \, d\tau, \int_0^1 U_c(\tau)^2 \, d\tau \right)$$

implies that the translation from B to U_c requires the inverse-mapping substitution $(r, s) \leftrightarrow (r + cs, s)$. Accordingly, Girsanov's transformation theorem gives

$$h_c(r, s) = \exp\left[cr + \frac{c^2}{2} s \right] h_0(r + cs, s).$$

Substituting for the known functional form $h_0(\cdot, \cdot)$ gives the stated result. ∎

Proof of Corollary 1. The proof follows from Theorem 2, by the methods developed in Abadir (1992,1995). ∎

LEMMA 1.   *For* $\operatorname{Re} p_1 \in \mathcal{R}_+$ *and* $\operatorname{Re} p_2 \in \mathcal{R}_+$,

$$\int_{\mathcal{R}_+} s^{\nu-1} (s + a)^\mu \exp\left[-p_1 s - \frac{p_2}{s} + p_3 \sqrt{s} + \frac{p_4}{\sqrt{s}} \right] ds$$

$$\sim \sqrt{\frac{\pi}{p_2}} \left(\frac{p_2}{p_1} \right)^{\frac{\nu}{2} + \frac{1}{4}} \left(\sqrt{\frac{p_2}{p_1}} + a \right)^\mu \exp\left[-2\sqrt{p_1 p_2} + p_3 \left(\frac{p_2}{p_1} \right)^{\frac{1}{4}} + p_4 \left(\frac{p_1}{p_2} \right)^{\frac{1}{4}} \right].$$

Proof of Lemma 1. Applying the asymptotic expansion of the Bessel K function to an integral in Oberhettinger and Badii (1973, p.41), we have

$$\int_{\mathcal{R}_+} s^{\nu-1} \exp\left[-p_1 s - \frac{p_2}{s} \right] ds \sim \sqrt{\frac{\pi}{p_2}} \left(\frac{p_2}{p_1} \right)^{\frac{\nu}{2} + \frac{1}{4}} \exp\left[-2\sqrt{p_1 p_2} \right].$$

We can therefore write

$$\int_{\mathcal{R}_+} s^{\nu-1} (s+a)^\mu \exp\left[-p_1 s - \frac{p_2}{s} + p_3\sqrt{s} + \frac{p_4}{\sqrt{s}}\right] ds$$

$$= \sum_{i=0}^{\infty} \binom{\mu}{i} a^{\mu-i} \sum_{j=0}^{\infty} \frac{p_3^j}{j!} \sum_{k=0}^{\infty} \frac{p_4^k}{k!} \int_{\mathcal{R}_+} s^{i+\frac{j-k}{2}+\nu-1} \exp\left[-p_1 s - \frac{p_2}{s}\right] ds$$

$$\sim \sum_{i=0}^{\infty} \binom{\mu}{i} a^{\mu-i} \sum_{j=0}^{\infty} \frac{p_3^j}{j!} \sum_{k=0}^{\infty} \frac{p_4^k}{k!} \sqrt{\frac{\pi}{p_2}} \left(\frac{p_2}{p_1}\right)^{\frac{i}{2}+\frac{j-k}{4}+\frac{\nu}{2}+\frac{1}{4}} \exp\left[-2\sqrt{p_1 p_2}\right]$$

$$= \sqrt{\frac{\pi}{p_2}} \left(\frac{p_2}{p_1}\right)^{\frac{\nu}{2}+\frac{1}{4}} \left(\sqrt{\frac{p_2}{p_1}} + a\right)^\mu \exp\left[-2\sqrt{p_1 p_2} + p_3\left(\frac{p_2}{p_1}\right)^{\frac{1}{4}} + p_4\left(\frac{p_1}{p_2}\right)^{\frac{1}{4}}\right]$$

as required. ■

Proof of Corollary 2. To obtain the required functions, we need

$$\frac{d\Pi(\kappa)}{da_1} = \frac{\exp\left[-\frac{c}{2}\right]}{\sqrt{\pi}} \sum_{j=0}^{\infty} \binom{j-\frac{1}{2}}{j} \sum_{\ell=0}^{j} \binom{j}{\ell} \frac{(-2)^\ell}{\Gamma\left(\ell+\frac{1}{2}\right)}$$

$$\int\int_{\mathcal{R}_+^2} \frac{r_+^{\ell-\frac{1}{2}}}{s^{\frac{\ell}{2}+\frac{5}{4}}} D_{\ell+\frac{3}{2}}^+ \left(\frac{r_+ + 2j + \frac{1}{2}}{\sqrt{s}}\right) \phi\left(a_1 + a_2 \frac{r_+ - cs - \frac{1}{2}}{\sqrt{s}} + a_3 c\sqrt{s}\right)$$

$$\exp\left[cr_+ - \left(2j+\frac{1}{2}\right)\frac{r_+}{s} - \frac{r_+^2}{2s}\right] dr_+ \exp\left[-\left(2j^2+j+\frac{1}{8}\right)\frac{1}{s} - \frac{c^2 s}{2}\right] ds$$

$$= \frac{\exp\left[-\frac{c}{2}\right]}{\pi\sqrt{2}} \sum_{j=0}^{\infty} \binom{j-\frac{1}{2}}{j} \sum_{\ell=0}^{j} \binom{j}{\ell} \frac{(-2)^\ell}{\Gamma\left(\ell+\frac{1}{2}\right)}$$

$$\int\int_{\mathcal{R}_+^2} r_+^{\ell-\frac{1}{2}} D_{\ell+\frac{3}{2}}^- \left(\frac{r_+ + 2j + \frac{1}{2}}{\sqrt{s}}\right) \exp\left[-\frac{1}{2}\left((a_3-a_2)c\sqrt{s} - \frac{a_2}{2\sqrt{s}} + a_1 + \frac{a_2 r_+}{\sqrt{s}}\right)^2\right]$$

$$\exp\left[cr_+\right] dr_+ \exp\left[-\frac{c^2 s}{2}\right] \frac{ds}{s^{\frac{\ell}{2}+\frac{5}{4}}},$$

where we have used a_1, a_2, a_3 (none of which contains c asymptotically) to simplify the notation, and rewritten D^+ in terms of $D_\nu^-(z) \equiv \exp[-z^2/2] D_\nu^+(z)$. For large $c \ll 0$, Corollary 1 shows that the leading term is the one for which $j = \ell = 0$. Furthermore, (28) shows that the transformations $r_+ \leftrightarrow r_+/\sqrt{|c|}$ and $s \leftrightarrow s/|c|$ would stabilize the integral as c changes. Applying these transformations, we can expand D^+ asymptotically by

$$D_{-\frac{1}{2}}^+(z) \equiv \frac{1}{\phi(z)\sqrt{2\pi}} D_{-\frac{1}{2}}^-(z) \sim \begin{cases} \frac{1}{\phi(z)\sqrt{-\pi z}}, & \text{as } z \to -\infty \\ \frac{1}{\sqrt{z}}, & \text{as } z \to \infty, \end{cases}$$

to get

$$\frac{d\Pi(\kappa)}{da_1} \sim \frac{\exp\left[-\frac{c}{2}\right]}{\sqrt{2\pi^3}} \int\int_{\mathcal{R}_+^2} \exp\left[-r_+\sqrt{|c|}\right] D_{\frac{3}{2}}^- \left(\frac{r_+ + \frac{1}{2}\sqrt{|c|}}{\sqrt{s}}\right)$$

$$\exp\left[-\frac{1}{2}\left((a_2-a_3)\sqrt{|c|s} - \frac{a_2}{2}\sqrt{\frac{|c|}{s}} + a_1 + \frac{a_2 r_+}{\sqrt{s}}\right)^2\right] \frac{dr_+}{\sqrt{r_+}} \exp\left[-\frac{|c|s}{2}\right] \frac{ds}{s^{\frac{5}{4}}}$$

$$\sim \frac{|c|^{\frac{3}{4}}\exp\left[-\frac{c}{2}\right]}{4\pi^{\frac{3}{2}}}\int\!\!\int_{\mathcal{R}_+^2}\exp\left[-r_+\sqrt{|c|}-\frac{\left(r_++\frac{1}{2}\sqrt{|c|}\right)^2}{2s}\right]$$

$$\exp\left[-\frac{1}{2}\left((a_2-a_3)\sqrt{|c|\,s}-\frac{a_2}{2}\sqrt{\frac{|c|}{s}}+a_1+\frac{a_2r_+}{\sqrt{s}}\right)^2\right]\frac{dr_+}{\sqrt{r_+}}\exp\left[-\frac{|c|\,s}{2}\right]\frac{ds}{s^2}$$

$$=\frac{|c|^{\frac{3}{4}}\exp\left[\left(1+a_2^2-a_2a_3\right)\frac{|c|}{2}-\frac{a_1^2}{2}\right]}{4\pi\left(1+a_2^2\right)^{\frac{1}{4}}}$$

$$\int_{\mathcal{R}_+}\exp\left[-\frac{\left(1+(a_2-a_3)^2\right)|c|\,s}{2}-\frac{\left(1+a_2^2\right)|c|}{8s}+a_1\left(a_3-a_2\right)\sqrt{|c|\,s}+\frac{a_1a_2}{2}\sqrt{\frac{|c|}{s}}\right]$$

$$D_{-\frac{1}{2}}^+\left(\frac{\left(1+a_2^2-a_2a_3\right)\sqrt{|c|\,s}+\frac{(1-a_2^2)\sqrt{|c|}}{2\sqrt{s}}+a_1a_2}{\sqrt{1+a_2^2}}\right)\frac{ds}{s^{\frac{7}{4}}}$$

from an integral in Oberhettinger and Badii (1973, p.43). We now have a new D^+ whose argument is large but not necessarily positive as $c \to -\infty$. Expanding D^+ asymptotically then integrating by means of our lemma,

$$\frac{d\Pi(\kappa)}{da_1}$$

$$\sim \mathrm{Re}\left\{\frac{\sqrt{|c|}\exp\left[\frac{1+a_2^2-a_2a_3}{1+a_2^2}|c|-\frac{a_1^2}{2(1+a_2^2)}\right]}{\pi\sqrt{8\left(a_2a_3-a_2^2-1\right)}}\int_{\mathcal{R}_+}s^{-\frac{3}{2}}\left(s+\frac{1-a_2^2}{2\left(1+a_2^2-a_2a_3\right)}\right)^{-\frac{1}{2}}\right.$$

$$\exp\left[-\frac{a_3^2\,|c|\,s}{2\left(1+a_2^2\right)}-\frac{a_2^2\,|c|}{2\left(1+a_2^2\right)s}+\frac{a_1a_3\sqrt{|c|\,s}}{1+a_2^2}+\frac{a_1a_2\sqrt{|c|}}{\left(1+a_2^2\right)\sqrt{s}}\right]ds$$

$$+\frac{\sqrt{|c|}\exp\left[\left(1+a_2^2-a_2a_3\right)\frac{|c|}{2}-\frac{a_1^2}{2}\right]}{4\pi\sqrt{1+a_2^2-a_2a_3}}\int_{\mathcal{R}_+}s^{-\frac{3}{2}}\left(s+\frac{1-a_2^2}{2\left(1+a_2^2-a_2a_3\right)}\right)^{-\frac{1}{2}}$$

$$\left.\exp\left[-\frac{\left(1+(a_2-a_3)^2\right)|c|\,s}{2}-\frac{\left(1+a_2^2\right)|c|}{8s}+a_1\left(a_3-a_2\right)\sqrt{|c|\,s}+\frac{a_1a_2}{2}\sqrt{\frac{|c|}{s}}\right]ds\right\}$$

$$\sim \mathrm{Re}\left\{\frac{\sqrt{1+a_2^2}\exp\left[\frac{1+a_2^2-(1+\mathrm{sgn}\,(a_2a_3))a_2a_3}{1+a_2^2}|c|-\frac{a_1^2}{2(1+a_2^2)}+\frac{a_1(\mathrm{sgn}\,(a_2)+\mathrm{sgn}\,(a_3))\sqrt{|a_2a_3c|}}{1+a_2^2}\right]}{\sqrt{2\pi a_2^2\left(2\left|\frac{a_2}{a_3}\right|\left(a_2a_3-a_2^2-1\right)+a_2^2-1\right)}}\right.$$

$$+\frac{\exp\left[\left(1+a_2^2-a_2a_3-\sqrt{\left(1+(a_2-a_3)^2\right)\left(1+a_2^2\right)}\right)\frac{|c|}{2}\right]}{\sqrt{\pi\left(1+a_2^2\right)\left(\left(1+a_2^2-a_2a_3\right)\sqrt{\frac{1+a_2^2}{1+(a_2-a_3)^2}}+1-a_2^2\right)}}$$

$$\exp\left[-\frac{a_1^2}{2}+\left((a_3-a_2)\left(\frac{1+a_2^2}{4\left(1+(a_2-a_3)^2\right)}\right)^{\frac{1}{4}}+\frac{a_2}{2}\left(\frac{4\left(1+(a_2-a_3)^2\right)}{1+a_2^2}\right)^{\frac{1}{4}}\right)a_1\sqrt{|c|}\right]\right\}.$$

Since the approximation is for the tail area of the density $d\Pi\left(\kappa\right)/da_1$, we need to integrate over the tail to obtain $1-\Pi\left(\kappa\right)$. Integrating with respect to a_1 by means of

$$\int_{a_1}^{\infty}\frac{1}{\sqrt{2\pi v}}\exp\left[\frac{wu}{v}-\frac{w^2}{2v}\right]dw=\exp\left[\frac{u^2}{2v}\right]\int_{a_1}^{\infty}\frac{1}{\sqrt{2\pi v}}\exp\left[-\frac{(w-u)^2}{2v}\right]dw$$

$$=\exp\left[\frac{u^2}{2v}\right]\Phi\left(\frac{u-a_1}{\sqrt{v}}\right),$$

we get

$$1-\Pi\left(\kappa\right)$$

$$\sim\operatorname{Re}\left\{\frac{\left(1+a_2^2\right)\exp\left[|c|\right]}{\sqrt{a_2^2\left(2\left|\frac{a_2}{a_3}\right|\left(a_2a_3-a_2^2-1\right)+a_2^2-1\right)}}\Phi\left(\frac{\left(\operatorname{sgn}\left(a_2\right)+\operatorname{sgn}\left(a_3\right)\right)\sqrt{|a_2a_3c|}-a_1}{\sqrt{1+a_2^2}}\right)\right.$$

$$+\frac{\sqrt{2}\exp\left[\left(1-\left(1+\frac{a_2^2}{2}\right)\sqrt{\frac{1+(a_2-a_3)^2}{1+a_2^2}}+(a_3-a_2)^2\sqrt{\frac{1+a_2^2}{4\left(1+(a_2-a_3)^2\right)}}\right)\frac{|c|}{2}\right]}{\sqrt{\left(1+a_2^2\right)\left(\left(1+a_2^2-a_2a_3\right)\sqrt{\frac{1+a_2^2}{1+(a_2-a_3)^2}}+1-a_2^2\right)}}$$

$$\left.\Phi\left(\left((a_3-a_2)\left(\frac{1+a_2^2}{4\left(1+(a_2-a_3)^2\right)}\right)^{\frac{1}{4}}+\frac{a_2}{2}\left(\frac{4\left(1+(a_2-a_3)^2\right)}{1+a_2^2}\right)^{\frac{1}{4}}\right)\sqrt{|c|}-a_1\right)\right\}.$$

Substituting for

$$a_1=\frac{\kappa}{\sqrt{1-\rho_q^2}}$$

$$a_2=-\frac{\rho_q}{\sqrt{1-\rho_q^2}}$$

$$a_3=-\frac{1}{\tilde{\sigma}_q\sqrt{1-\rho_q^2}}<0$$

gives the stated result for the Studentized t statistic $(\Pi=\Pi_t)$, while

$$a_1=0$$

$$a_2=-\frac{\rho_q}{\sqrt{1-\rho_q^2}}$$

$$a_3=\frac{\frac{\kappa}{c}-1}{\tilde{\sigma}_q\sqrt{1-\rho_q^2}}$$

gives the result for the autocorrelation coefficient $(\Pi=\Pi_\phi)$.   ■

REFERENCES

Abadir, K. (1992). A distribution generating equation for unit-root statistics. *Oxford Bulletin of Economics and Statistics* **54** 305–323.

Abadir, K. (1993). Expansions for some confluent hypergeometric functions. *J. Phys. A: Math. Gen.* **26** 4059–4066. [Corrigendum for printing error (1993) 7663].

Abadir, K. (1995). The joint density of two functionals of a Brownian motion. *Math. Methods of Statist.* **4** 449–462. [Correction and generalization (1996) 124,505].

Abadir, K. (1999). An introduction to hypergeometric functions for economists. *Econometric Reviews* **18** 287–330.

Cameron, R. and Martin, W. (1944). Transormation of Wiener integrals under translations. *Ann. Math.* **45** 386–396.

Cameron, R. and Martin, W. (1945). Transormation of Wiener integrals under a general class of linear transformations. *Trans. Amer. Math. Soc.* **58** 148–219.

Chan, N. and Wei, C. (1987). Asymptotic inference for nearly nonstationary ar(1) processes. *Ann. Statist.* **15** 1050–1063.

Cox, D. and Llatas, I. (1991). Maximum lieklihood type estimation for nearly nonstationary autoregressive time series. *Ann. Statist.* **19** 1109–1128.

Elliott, G., Rothenberg, T. and Stock, J. (1996). Efficient tests for an autoregressive unit root. *Econometrica* **64** 813–836.

Erdélyi, A. (1953). *Higher transcedental functions*, Volumes 1–2. McGraw-Hill, New York.

Fuller, W. (1976). *Introduction to Statistical Time Series*. Wiley, New York.

Girsanov, I. (1960). On transformations of one class of random processes with the help of absolutely continuous substitution of the measure. *Theory Probab. Appl.* **3** 314–330.

Lehmann, E. (1986). *Testing statistical hypotheses* (2nd ed.). Wiley, New York.

Lucas, A. (1995). Unit root tests based on M estimators. *Econometric Theory* **11** 331–346.

Oberhettinger, F. and Badii, L. (1973). *Tables of Laplace transforms*. Springer-Verlag, Berlin.

Perron, P. (1989). The calculation of the limiting distribution of the least-squares estimator in a near-integrated model. *Econometric Theory* **5** 241–255.

Perron, P. (1991). A continuous time approximation to the unstable first-order autoregressive process: the case without an intercept. *Econometrica* **59** 211–236.

Phillips, P. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika* **74** 535–547.

Rothenberg, T. and Stock, J. (1997). Inference in a nearly integrated autoregressive model with nonnormal innovations. *J. Econometrics* **80** 269–286.

Rustagi, J. (1976). *Variational methods in Statistics*. Academic, New York.

White, J. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.* **29** 1188–1197.

DEPT. OF MATHEMATICS AND
DEPT. OF ECONOMICS AND RELATED
  STUDIES
UNIVERSITY OF YORK
HESLINGTON
YO10 5DD UK
EMAIL: KMA4@YORK.AC.UK.

DEPT. FINANCE AND TINBERGEN INSTITUTE
ECO/FIN, VRIJE UNIVERSITEIT
DE BOELELAAN 1105
1081HV AMSTERDAM
THE NETHERLANDS
EMAIL: ALUCAS@ECON.VU.NL.