A. Introduction

The purpose of this paper is to establish a notation and derive certain distributional results needed in subsequent work. The topic of this research is statistical time series analysis in the time domain, especially autoregressive-moving average (ARMA) models. The particular goal is a structured Bayesian approach to ARMA time series models. In this paper, a portion of the entire work, a notation is established for the analysis of these models. Then the results of the classical statistical approach, maximum likelihood are obtained. The structure of the prior distributions to be used in the Bayesian analysis is next described, followed by some derivations of the necessary conditional and/or posterior distributions. The proofs of these derivations are given in Appendix G.

B. Notation and Definitions

Consider the autoregressive-moving average (ARMA) process $\{z_t\}$ of order (p,q) described by (B.1):

$$\text{(B.1)} \quad (z_t-\mu) - \phi_1(z_{t-1}-\mu) - \cdots - \phi_p(z_{t-p}-\mu) = e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}$$

where the e_t's are iid normal random variables each with mean zero and variance σ_e^2. A finite segment of this process is observed: $z = (z_1, \ldots, z_N)^T$ which has an N dimensional multivariate normal distribution with mean vector $\mu 1_N$ and covariance matrix $\sigma_e^2 A_N$. The matrix A_N is described by

$$\text{(B.2)} \quad (A_N)_{ij} = \text{Cov}(z_i, z_j) = \sigma(i-j) = \sigma(|i-j|)$$

where the covariance function σ characterizes this time series process. By taking variances of both sides of (B.1), the covariance function σ can be determined from

$$\text{(B.3)} \quad \text{Cov}\left(\sum_{i=0}^{p} \phi_i z_{t-i}, \sum_{j=0}^{p} \phi_j z_{t+s-j}\right) = \sum_{i=0}^{p}\sum_{j=0}^{p} \phi_i \phi_j \sigma(s+i-j)\sigma_e^2$$

$$= \sum_{j=0}^{q} \theta_j \theta_{j+s} \sigma_e^2$$

for any integer $s \geq 0$, where $\phi_0 = \theta_0 = -1$ and $\phi_i = \theta_j = 0$ for $i > p$ or $j > q$. (See Anderson (1971, p. 237); McLeod (1975) gives an algorithm for computing σ.)

Since forecasting is of primary interest, the distribution of the future n observations, that is, $z_F = (z_{N+1}, \ldots, z_{N+n})^T$, conditional on the observed z, has a multivariate normal distribution with mean $\mu 1_n + A_{21} A_N^{-1} (z - \mu 1_N)$ and covariance matrix $\sigma_e^2 A_{nN}^*$ where

$$A_{N+n} = \begin{pmatrix} A_N & A_{12} \\ A_{21} & A_n \end{pmatrix} \begin{matrix} \} \; N \\ \} \; n \end{matrix}$$

and $A_{nN}^* = A_n - A_{21} A_N^{-1} A_{12}$. In the notation to be used here, this is written

(B.4) $$(z_F | \mu, \sigma_e^2, z) \sim N(\mu 1_n + A_{21} A_N^{-1} (z - \mu 1_N),\ \sigma_e^2 A_{nN}^*) \ .$$

Notice that A_{N+n} is a function of $\psi = (\phi_1, \ldots, \phi_p,\ \theta_1, \ldots, \theta_q)^T$.

C. Maximum Likelihood

The analysis for applying the maximum likelihood approach is rather straightforward when the model $M = (p,q)$ is assumed to be known. The density of the observations is given by

$$\text{(C.1)} \quad p(z|\mu,\sigma_e^2,\psi) = (2\pi)^{-N/2}|\sigma_e^2 A_N|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_e^2}(z-\mu 1_N)^T A_N^{-1}(z-\mu 1_N)\right\}$$

By taking the natural logarithm of both sides above, the log-likelihood function is written

$$\text{(C.2)} \quad \ell(\mu,\sigma_e^2,\psi) = C_1 - \frac{N}{2}\ln\sigma_e^2 - \tfrac{1}{2}\ln|A_N| - \frac{1}{2\sigma_e^2}(z-\mu 1_N)^T A_N^{-1}(z-\mu 1_N)$$

where C_1 is a constant (as are later C_i's). First, by setting $\partial\ell(\mu,\sigma_e^2,\psi)/\partial\mu$ to zero, the maximum likelihood estimator (MLE) for μ can be found to be

$$\text{(C.3)} \qquad \hat{\mu} = 1_N^T A_N^{-1} z / 1_N^T A_N^{-1} 1_N \ .$$

The MLE of σ_e^2 can then be found by solving $\partial\ell(\hat{\mu},\sigma_e^2,\psi)/\partial\sigma_e^2 = 0$, which yields

$$\text{(C.4)} \qquad \hat{\sigma}_e^2 = \frac{1}{N}Q(\psi) = \frac{1}{N}(z-\hat{\mu}1_N)^T A_N^{-1}(z-\hat{\mu}1_N) \ .$$

The concentrated log-likelihood function can now be written

$$\text{(C.5)} \qquad \ell^*(\psi) = C_2 - N/2 \ln Q(\psi) - \tfrac{1}{2}\ln|A_N| \ .$$

The MLE of the remaining parameters, ψ, is that vector which maximizes $\ell^*(\psi)$, which must be found numerically. Note that for identifiability purposes, the values of the parameter vector ψ must be restricted to that region which insures the process $\{z_t\}$ to be stationary and invertible.

The maximum likelihood approach to forecasting requires the examination of the statistical properties of maximum likelihood estimates under standard regularity conditions. The great appeal of maximum likelihood methods lies in the performance of MLE's when large samples (i.e., large N) are encountered. Subject to mild regularity conditions, MLE's are consistent and asymptotically normally distributed about the true parameter values.

Let the entire parameter vector be written as $\eta = (\mu,\sigma_e^2,\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q)^T$; $\hat{\eta} = (\hat{\mu},\hat{\sigma}_e^2,\hat{\psi}^T)^T$ denotes the MLE for η, that is, that vector η which maximizes $\ell(\eta)$ from (C.2); and let η^* denote the true values of the unknown parameters. The asymptotic statement just mentioned can be written more practically as

$$\text{(C.6)} \qquad \hat{\eta} \underset{\text{(approx)}}{\sim} N(\eta^*, D_N(\eta^*))$$

where $D_N(\eta^*)$ is the asymptotic covariance matrix of $\hat{\eta}$, based on N observations z. Theoretically, $D_N(\eta^*)$ can be obtained from Fisher's Information matrix (Theil (1971, p. 192 ff) explains the simple vector case) but this is not feasible here. However, a consistent estimate $D_N(\hat{\eta})$ can be obtained by taking numerical second partial derivatives of $\ell(\eta)$ at the MLE $\hat{\eta}$.

Since the problem at hand is forecasting, the forecasting vector following the maximum likelihood approach is

$$\hat{z}_F = z_F(\hat{\eta}, z) = \hat{\mu}1_n + A_{21}(\hat{\psi})(A_N(\hat{\psi}))^{-1}(z - \hat{\mu}1_N) \tag{C.7}$$

which is found be replacing the unknown parameters in the mean given in (B.4). Since the elements of η are unknown, the "covariance matrix" needed to obtain standard errors for $\hat{z}_F$ must be augmented. An asymptotic distribtuion result similar to (C.6) can be written

$$(z_F(\hat{\eta}, z) \mid z) \underset{(\text{approx})}{\sim} N(z_F(\eta^*, z),\ D_{F,N}(\eta^*)) \tag{C.8}$$

where

$$D_{F,N}(\eta^*) = \sigma_e^2 A_{nN}^*(\psi) + \left[\frac{\partial z_F(\eta, z)}{\partial \eta^T}\right]\Bigg|_{\eta=\eta^*} D_N(\eta^*) \left[\frac{\partial z_F(\eta, z)}{\partial \eta^T}\right]\Bigg|_{\eta=\eta^*}^T .$$

The covariance matrix above in (C.8) can be consistently estimated by

$$\hat{D}_{F,N}(\hat{\eta}) = \hat{\sigma}_e^2 A_{nN}^*(\hat{\psi}) + \left[\frac{\partial z_F(\eta, z)}{\partial \eta^T}\right]\Bigg|_{\eta=\hat{\eta}} \hat{D}_N(\hat{\eta}) \left[\frac{\partial z_F(\eta, z)}{\partial \eta^T}\right]\Bigg|_{\eta=\hat{\eta}}^T . \tag{C.9}$$

D. Bayesian Analysis

For the Bayesian analysis in this paper, the prior distribution takes on a highly structured form, described below, in order that the analysis flows smoothly. It is hoped that the structure is sufficiently flexible to admit the expression of a great variety of opinions.

The prior distribution is expressed in many levels of increasing fineness. At the coarsest level, the prior probability distribution admits probability to the discrete set of models M described by the order of the ARMA process, (p,q). That is, $p(m)$, $m = 1,2,3,\ldots$ are the probabilities for which

$$p(m) = Pr(M{=}m) = Pr(z \text{ arises from an ARMA } (p_m,q_m) \text{ process})$$

where, by the convention chosen here, $m = (q_m+1) + (p_m+q_m)(p_m+q_m+1)/2.$[1/] A density function (with respect to Lebesgue measure on $R^{p_m + q_m}$), $\pi(\psi|M{=}m)$, describes the prior distribution of the ARMA parameters, $\psi = (\phi_1,\ldots,\phi_p, \theta_1,\ldots,\theta_q)^T$ of the ARMA (p,q) process yielding the observations, z. The support of $\pi(\psi|M{=}m)$ is restricted to that subset of $R^{p_m+q_m}$ for which the ARMA $(p_m q_m)$ process described by (B.1) is stationary and invertible.[2/]

[1/] Hence $m = 1$ for $(0,0)$, $m = 2$ for $(1,0),\ldots$, $m = 6$ for $(0,2)$.

[2/] This is not a constraint but a manner in which identifiability is established. See Box and Jenkins (1976, pp. 195-198); they use the term "model multiplicity." The identifiability regions (or stationary and invertibility regions) become complicated for higher order models. For the Bayesian approach, this problem is easily surmountable. A method for integrating over these regions is explained in Monahan (1977).

The remaining parameters are now μ, the mean of the process, and σ_e^2, the disturbance variance, which will hereafter (in the Bayesian analysis) be expressed as r, the disturbance precision, related by $r \equiv 1/\sigma_e^2$. The prior distribution of μ given the precision r (and, implied, M and ψ) is normal with mean γ and precision τr, or[3/]

(D.1) $$(\mu|r) \sim N(\gamma,(\tau r)^{-1})$$

Finally, r has a gamma distribution with shape parameter α and scale parameter β:

(D.2) $$(r) \sim \text{gamma } (\alpha,\beta)$$

Notice that the parameters of these last levels of prior distributions, γ,τ,α,β, may depend on ψ and M.

Since the distribution of the observations z has already been stated (in Section B),

(D.3) $$(z|M,\psi,\mu,r) \sim N(\mu 1_N, r^{-1}A_N) \ ,$$

certain posterior distributions follow. Proofs of the analyses are given in Appendix G. To summarize, understanding that all are for given ψ and M,

(D.4a) $$(\mu|z,r) \sim N\left(\frac{\gamma\tau + 1_N^T A_N^{-1} z}{\tau + 1_N^T A_N^{-1} 1_N}\ ,\ r^{-1}(\tau + 1_N^T A_N^{-1} 1_N)^{-1}\right)$$

(D.4b) $$(\mu|z) \sim t_{2\alpha+N}\left(\gamma^*\ ,\ \left(\frac{2\alpha+N}{2\beta^*}\right)\ \tau^*\right)$$

3/ Definitions of the distributional notation are given in Appendix E.

(D.4c) $(r|z) \sim \text{gamma}(\alpha + \frac{N}{2}\ ,\ \beta + \tfrac{1}{2}(z-\gamma 1_N)^T(A_N^{-1} + \tau^{-1}1_N 1_N^T)^{-1}(z-\gamma 1_N))$

(D.4d) $(z) \sim t_{N,2\alpha}(\gamma 1_N,\ \frac{\alpha}{\beta}\ (A_N + \tau^{-1}1_N 1_N^T)^{-1})$

(D.4e) $(z_F|z,r) \sim N(\gamma^* a + b,\ r^{-1}(A_{nN}^* + \tau^{-*}aa^T))$

(D.4f) $(z_F|z) \sim t_{n,2\alpha+N}\left(\gamma^* a + b,\ \left(\frac{2\alpha+N}{2\beta^*}\right)\ (A_{nN}^* + \tau^{-*}aa^T)^{-1}\right)$

where the following notation has been used:

$$\gamma^* = \frac{\gamma\tau + 1_N^T A_N^{-1} 1_N}{\tau + 1_N^T A_N^{-1} 1_N} \quad \text{as in (D.4a)}$$

$$a = 1_n - A_{21}A_N^{-1}1_N$$

$$b = A_{21}A_N^{-1}z$$

$$\tau^* = \tau + 1_N^T A_N^{-1} 1_N = (\tau^{-*})^{-1}$$

$$\beta^* = \beta + (z-\gamma 1_N)^T A_N^{-1}(z-\gamma 1_N) \quad \text{as in (D.4c)}\ .$$

The remainder of the Bayesian analysis now follows easily:

(D.5) $\pi(\psi|M,z) = \pi(\psi|M)p(z|\psi,M)/p(z|M)$

where

(D.6) $p(z|M) = \int \pi(\psi|M)p(z|\psi,M)d\psi$

(D.7) $p(M|z) = p(M)p(z|M)/\sum_j p(j)p(z|j)\ .$

And the posterior distribution of the forecasts follows from

$$\text{(D.8)} \qquad p(z_F|z) = \sum_j \int p(z_F|z,\psi,M=j)\pi(\psi|z,M=j)p(j|z)d\psi$$

where $p(z_F|z,\psi,M)$ is found from (D.4f).

In the event that μ and/or r are known, the preceding analysis must be modified somewhat. If both μ and r are known, all of the analysis in the Appendix can be bypassed; the results listed in (D.4) are unnecessary; the only distributions needed are that of z and z_F: (D.3) and (B.4). If, however, r alone is known, (D.4a) is unchanged, (D.4b) and (D.4f) are dropped, and (D.4d) is replaced by

$$\text{(D.4d}^*\text{)} \qquad (z|r,\psi,M) \sim N(\gamma 1_N,\ r^{-1}(A_N + \tau^{-1}1_N 1_N^T)) \ .$$

If, on the other hand, μ alone is known, the necessary results can be obtained by taking $\tau \to \infty$ so that $\mu = \gamma$ with probability one.

$$\text{(D.4c}^{**}\text{)} \qquad (r|z) \sim \text{gamma}(\alpha + \frac{N}{2}\ ,\ \beta + (z-\mu 1_N)^T A_N^{-1}(z-\mu 1_N))$$

$$\text{(D.4d}^{**}\text{)} \qquad (z) \sim t_{N,2\alpha}(\mu 1_N,\ \frac{\alpha}{\beta} A_N^{-1})$$

$$\text{(D.4e}^{**}\text{)} \qquad (z_D|z,r) \sim (z_F|z,\mu,r) \quad \text{see (B.4)}$$

$$\text{(D.4f}^{**}\text{)} \qquad (z_F|z) \sim t_{N,2\alpha+N}\left(\mu a+b\ ,\ \left(\frac{2\alpha+N}{2\beta^{**}}\right)(A^*_{nN})^{-1}\right)$$

where β^{**} is the scale parameter in (D.4c**).

Appendix E: Distributional Definitions

A random variable X said to have the (k-dimensional) multivariate normal distribution with mean vector μ and covariance matrix A is denoted by

$$(X) \sim N(\mu, A)$$

where X and μ are k by 1 vectors and A is k by k and positive definite. The probability density function of such an X is

$$\text{(E.1)} \qquad p(x) = (2\pi)^{-k/2}|A|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(x-\mu)^T A^{-1}(x-\mu)\}.$$

If $k = 1$, X (now a scalar) has a univariate normal distribution. The same notation is used to write this; A is now a scalar, say σ^2, is called the variance, and must be positive. The density function has a simpler form:

$$p(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}.$$

Also, A^{-1} is sometimes referred to as the precision matrix of the multivariate normal; $1/\sigma^2$, the precision of the univariate normal.

A random variable X said to have a gamma distribution with shape parameter α and scale parameter β is denoted by

$$(X) \sim \text{gamma}(\alpha, \beta)$$

where both α and β are positive. The probability density function of such an X is

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad \text{for} \quad x > 0\,.$$

A random variable X said to have a (k-dimensional) <u>multivariate t-distribution</u> with n degrees of freedom, location vector μ and precision matrix B (see definition in DeGroot (1970, pp. 59-62) is described by

$$(X) \sim t_{k,n}(\mu, B)$$

where X and μ are k by 1 and B is a k by k positive definite matrix. The probability density function of such an X is given by

$$p(x) = \frac{\Gamma\left(\frac{n+k}{2}\right) |B|^{\frac{1}{2}}}{\Gamma\left(\frac{n}{2}\right) (n\pi)^{k/2}} \left[1 + \frac{1}{n} (x-\mu)^T B(x-\mu)\right]^{-\frac{n+k}{2}}$$

If $k = 1$, X is now a scalar and is said to have the univariate (Student's) t distribution, written

$$(X) \sim t_n(\mu, b)$$

where μ and b are scalars and $b > 0$. The density function of such an X is given by

$$p(x) = \frac{\Gamma\left(\frac{n+1}{2}\right) b^{\frac{1}{2}}}{\Gamma\left(\frac{n}{2}\right) \sqrt{n\pi}} \left(1 + \frac{(x-\mu)^2}{nb}\right)^{-\frac{n+1}{2}} .$$

Appendix F. Lemmas

<u>Lemma 1</u>: $|I_n + ab^T| = (1 + b^T a)$

<u>Proof</u>: Partition $a^T = (a_1, a^*)$, $b^T = (b_1, b_*)$ and write

$$|I_n + ab^T| = \begin{vmatrix} 1 + a_1 b_1 & a_1 b_*^T \\ b_1 a_1 & I_{n-1} + a_* b_*^T \end{vmatrix} = \begin{vmatrix} 1 + a_1 b_1 & a_1 b_*^T \\ -\frac{1}{a_1} a_* & I_{n-1} \end{vmatrix}$$

This last expression is obtained by subtracting $\left(\frac{a_i}{a_1}\right)$ times the first row from the ith:

$$\text{first column:} \quad b_1 a_i - \frac{a_i}{a_1}(1 + a_1 b_1) = -\frac{a_i}{a_1}$$

$$\text{diagonal:} \quad 1 + b_i a_i - \frac{a_i}{a_1}(a_1 b_i) = 1$$

$$\text{others:} \quad a_i b_j - \frac{a_i}{a_1}(a_1 b_j) = 0 \ .$$

These operations do not change the determinant and the last determinant can be found by partitioning:

$$|I_n + ab^T| = |I_{n-1}^{-1}|\,|(1 + a_1 b_1) - (a_1 b_*^T)(I_{n-1}^{-1})(-\frac{1}{a_1} a_*)|$$

$$= (1 + a_1 b_1 + b_*^T a_*) = (1 + b^T a)$$

<u>Lemma 2</u>: If A is nonsingular, $|A + \tau^{-1} dd^T| = |A|(1 + \tau^{-1} d^T A^{-1} d)$.

<u>Proof</u>: $|A + \tau^{-1} dd^T| = |A|\,|I + \tau^{-1} A^{-1} dd^T| = |A|(1 + \tau^{-1} d^T A^{-1} d)$

<u>Lemma 3</u>. If A is positive definite and $\tau > 0$,

$$(A + \tau^{-1} dd^T)^{-1} = A^{-1} - A^{-1} dd^T A^{-1} / (\tau + d^T A^{-1} d) \ .$$

Proof: $(A + \tau^{-1}dd^T)(A^{-1} - A^{-1}dd^TA^{-1}/(\tau + d^TA^{-1}d))$

$$= I + \tau^{-1}dd^TA^{-1} + [-dd^TA^{-1} - dd^TA^{-1}(\tau^{-1}d^TA^{-1}d)]/(\tau + d^TA^{-1}d)$$

$$= I + \tau^{-1}dd^TA^{-1} + dd^TA^{-1}(-\tau^{-1}) = I$$

Lemma 4. Given a positive definite matrix A, vectors d and z, scalars $\tau \geq 0$ and γ, LHS = RHS below

$$\text{LHS} = \gamma^2\tau + z^TA^{-1}z - (\gamma\tau + z^TA^{-1}d)^2/(\tau + d^TA^{-1}d)$$

$$\text{RHS} = (z - \gamma d)^T(A^{-1}-A^{-1}dd^TA^{-1}/(\tau + d^TA^{-1}d))(z - \gamma d)$$

$$= (z - \gamma d)^T(A + \tau^{-1}dd^T)^{-1}(z - \gamma d) \ .$$

Proof: Subtract $z^TA^{-1}z$ from both LHS and RHS and multiply both by $(\tau + d^TA^{-1}d)$ to get, respectively, LHS* and RHS*:

$$\text{LHS}^* = \gamma^2\tau(\tau + d^TA^{-1}d) - (\gamma\tau + z^TA^{-1}d)^2$$

$$= \gamma^2\tau^2 + \gamma^2\tau d^TA^{-1}d - \gamma^2\tau^2 - 2\gamma\tau z^TA^{-1}d - (z^TA^{-1}d)^2$$

$$= \gamma^2\tau d^TA^{-1}d - 2\gamma\tau z^TA^{-1}d - (z^TA^{-1}d)^2$$

$$\text{RHS}^* = (-2\gamma z^TA^{-1}d + \gamma^2d^TA^{-1}d)(\tau + d^TA^{-1}d) - (z - \gamma d)^T(A^{-1}dd^TA^{-1})(z - \gamma d)$$

$$= -2\gamma\tau z^TA^{-1}d + \tau\gamma^2d^TA^{-1}d - 2\gamma z^TA^{-1}dd^TA^{-1}d + \gamma^2(d^TA^{-1}d)^2$$

$$- (z^TA^{-1}d)^2 + 2\gamma(z^TA^{-1}d)(d^TA^{-1}d) - \gamma^2(d^TA^{-1}d)^2$$

$$= \tau\gamma^2d^TA^{-1}d - 2\gamma\tau z^TA^{-1}d - (z^TA^{-1}d)^2$$

Appendix G. Theorems

NB: Subscripts have been dropped when unnecessary. Throughout this section z is N by 1 and z_F is n by 1.

<u>Theorem 1</u>. Given $(\mu|r) \sim N(\gamma,(\tau r)^{-1})$ and $(z|\mu,r) \sim N(\mu 1, r^{-1}A)$, then $(\mu|z,r) \sim N(\gamma^*, r^{-1}\tau^{-*})$ and $(z|r) \sim N(\gamma 1, r^{-1}(A+\tau^{-1}11^T)^{-1})$

<u>Proof</u>: $p(\mu|r)p(z|\mu,r) = p(\mu,z|r) = p(\mu|z,r)p(z|r)$

$$p(\mu,z|r) = (2\pi)^{-\frac{N+1}{2}} \tau^{\frac{1}{2}} r^{\frac{N+1}{2}} |A|^{-\frac{1}{2}} \exp\{-\tfrac{r}{2}[(\mu-\gamma)^2\tau + (z-\mu 1)^T A^{-1}(z-\mu 1)]\}$$

$$\begin{aligned}[] = {} & \mu^2\tau^* - 2\mu(\gamma\tau + z^T A^{-1}1) + (\gamma\tau + z^T A^{-1}1)^2/\tau^* \\ & + z^T A^{-1} z + \gamma^2\tau - (\gamma\tau + z^T A^{-1}1)^2/\tau^* \\ = {} & \tau^*(\mu - \gamma^*)^2 + (z - \gamma 1)^T(A^{-1} - \tau^{-*}A^{-1}11^T A^{-1})(z - \gamma 1) \\ = {} & \tau^*(\mu - \gamma^*)^2 + (z - \gamma 1)^T(A + \tau^{-1}11^T)^{-1}(z - \gamma 1) \quad \text{via Lemma 4.}\end{aligned}$$

$$p(\mu|z,r)p(z|r) = (2\pi)^{-\frac{1}{2}}(\tau^*)^{\frac{1}{2}} r^{\frac{1}{2}} \exp\{-\tfrac{r}{2}(\mu-\gamma^*)^2\tau^*\}$$
$$\times (2\pi)^{-\frac{N}{2}} r^{\frac{N}{2}} |A|^{-\frac{1}{2}}(\tau^*/\tau)^{-\frac{1}{2}} \exp\{-\tfrac{r}{2}(z-\gamma 1)^T(A^T+\tau^{-1}11^T)^1(z-\gamma 1)\} .$$

Notice: $|A + \tau^{-1}11^T| = |A|(1 + \tau^{-1}1^T A^{-1}1) = |A|(\tau^*/\tau)$.

Theorem 2. Given $(r) \sim \text{gamma}(\alpha,\beta)$ and $(z|r) \sim N(\gamma 1, r^{-1}(A+\tau^{-1}11^T))$ then $(r|z) \sim \text{gamma}(\alpha+\frac{N}{2}, \beta^*)$ and $(z) \sim t_{N,2\alpha}(\gamma 1, \frac{\alpha}{\beta}(A+\tau^{-1}11^T)^{-1})$ where $\beta^* = \beta + \frac{1}{2}(z-\gamma 1)^T(A+\tau^{-1}11^T)^{-1}(z-\gamma 1)$.

Proof.
$$p(z|r)p(r) = p(z,r) = p(r|z)p(z)$$

$$p(z,r) = (2\pi)^{-\frac{N}{2}} |A+\tau^{-1}11^T|^{-\frac{1}{2}} r^{\frac{N}{2}} \exp\{-\frac{r}{2}(z-\gamma 1)^T(A+\tau^{-1}11^T)^{-1}(z-\gamma 1)\}$$

$$\times \frac{\beta^\alpha}{\Gamma(\alpha)} r^{\alpha-1}e^{-\beta r}$$

$$= \frac{(\beta^*)^{\alpha+\frac{N}{2}}}{\Gamma\left(\frac{N}{2}+\alpha\right)} r^{\alpha+\frac{N}{2}-1} e^{-r\beta^*} \times \frac{\Gamma\left(\frac{N}{2}+\alpha\right)|A+\tau^{-1}11^T|^{-\frac{1}{2}}\left(\frac{\alpha}{\beta}\right)^{N/2}}{\Gamma(\alpha)(2\pi\alpha)^{\frac{N}{2}}\left[1+\frac{1}{2\beta}(z-\gamma 1)^T(A+\tau^{-1}11^T)^{-1}(z-\gamma 1)\right]^{(\alpha+\frac{N}{2})}}$$

Corollary. Given $(\mu|z,r) \sim N(\gamma^*, r^{-1}\tau^{-*})$ from Theorem 1, then $(\mu|z) \sim t_{2\alpha+N}(\gamma^*, \left(\frac{2\alpha+N}{2\beta^*}\right)\tau^*)$.

Proof.
$$p(\mu|z) = \int p(\mu|z,r)p(r|z)dr$$

$$= \int_0^\infty (2\pi)^{-\frac{1}{2}}(\tau^* r)^{\frac{1}{2}} e^{-\frac{r}{2}(\mu-\gamma^*)^2\tau^*} \cdot \frac{(\beta^*)^{\alpha+\frac{N}{2}}}{\Gamma\left(\alpha+\frac{N}{2}\right)} r^{\alpha+\frac{N}{2}-1} e^{-\beta^* r}\, dr$$

$$= (2\pi)^{-\frac{1}{2}}(\tau^*)^{\frac{1}{2}} \frac{(\beta^*)^{\alpha+\frac{N}{2}}}{\Gamma\left(\alpha+\frac{N}{2}\right)} \int_0^\infty r^{\alpha+\frac{N+1}{2}-1} e^{-r(\beta^*+\frac{1}{2}(\mu-\gamma^*)^2\tau^*)}\, dr$$

$$= (2\pi)^{-\frac{1}{2}}(\tau^*)^{\frac{1}{2}}(\beta^*)^{\alpha+\frac{N}{2}} \frac{\Gamma\left(\alpha+\frac{N+1}{2}\right)}{\Gamma\left(\alpha+\frac{N}{2}\right)} (\beta^*+\tfrac{1}{2}(\mu-\gamma^*)^2\tau^*)^{-(\alpha+\frac{N+1}{2})}$$

$$= \frac{\Gamma\left(\frac{2\alpha+N+1}{2}\right)(\tau^*)^{\frac{1}{2}}\left(\alpha+\frac{N}{2}\right)^{\frac{1}{2}}(\beta^*)^{-\frac{1}{2}}}{\Gamma\left(\frac{2\alpha+N}{2}\right)\left(\pi(2\alpha+N)\right)^{\frac{1}{2}}} \left[1+\frac{(\mu-\gamma^*)^2}{2\alpha+N}\left(\frac{\left(\alpha+\frac{N}{2}\right)}{\beta^*}\tau^*\right)\right]^{-\frac{2\alpha+N+1}{2}}$$

<u>Theorem 3</u>. Given $(z_F|z,\mu,r) \sim N(\mu 1_n + A_{21}A_N^{-1}(z-\gamma 1),\ r^{-1}A_{nN}^*)$ then $(z_F|z,r) \sim N(\gamma^* a + b,\ r^{-1}(A_{nN}^* + \tau^{-*}aa^T))$ where $\tau^{**} = \tau + 1^T A_N^{-1} 1 + a^T A^{-*} a$, $a = 1_n - A_{21}A_N^{-1}1_N$, $b = A_{21}A_N^{-1}z$ and $A^{-*} = (A_{nN}^*)^{-1}$.

<u>Proof</u>. $$p(z_F|z,r) = \int p(z_F|z,\mu,r)p(\mu|z,r)d\mu$$

where $p(\mu|z,r)$ is available from Theorem 1

$$p(z_F|z,\mu,r)p(\mu|z,r) = (2\pi)^{-\frac{N}{2}} |A_{nN}^*|^{-\frac{1}{2}} \cdot (2\pi)^{-\frac{1}{2}}(\tau^* r)^{\frac{1}{2}} \cdot$$

$$\times \exp\{-\tfrac{r}{2}[\tau^*(\mu-\gamma^*)^2 + (z_F-b-\mu a)^T A^{-*}(z_F-b-\mu a)]\}$$

$$\begin{aligned}[\,] &= \mu^2\tau^{**} - 2\mu(\gamma^*\tau^* \pm a^T A^{-*}(z_F-b)) \\ &\quad + (\gamma^*)^2\tau^* + (z_F-b)^T A^{-*}(z_F-b) \\ &= \tau^{**}(\mu-(\gamma^*\tau^* + aA^{-*}(z_F-b))/\tau^{**})^2 \\ &\quad + (\gamma^*)^2\tau^* + (z_F-b)^T A^{-*}(z_F-b) - (\gamma^*\tau^* + a^T A^{-*}(z_F-b))^2/\tau^{**}\end{aligned}$$

Integrating out μ and applying Lemma 4 while substituting $\tau = \tau^*$, $d = a$, $\gamma = \gamma^*$, $z = z_F - b$, $A = A^*$ will yield

$$p(z_F|z,r) = (2\pi)^{-\frac{n}{2}} r^{\frac{n}{2}} |A^*|^{-\frac{1}{2}} \left(\frac{\tau^*}{\tau^{**}}\right)^{\frac{1}{2}} \exp\left\{-\frac{r}{2}\left[(z_F-b-\gamma^* a)^T\left(A^* + \frac{aa^T}{\tau^*}\right)^{-1}(z_F-b-\gamma^* a)\right]\right\}$$

and notice $|A^* + aa^T/\tau^*| = |A^*|(\tau^{**}/\tau^*)$.

<u>Corollary</u>. $(z_F|z) \sim t_{n,2\alpha+N}\left(\gamma^* a+b,\ \left(\frac{2\alpha+N}{2\beta^*}\right)(A_{nN}^* + aa^T/\tau^*)^{-1}\right)$

Proof follows directly from $p(z_F|z) = \int p(z_F|z,r)p(r|z)\,dr$.

$$p(z_F|z) = \int_0^\infty p(z_F|z,r)p(r|z)dr$$

$$= \int_0^\infty (2\pi)^{-n/2}|A^*|^{-\frac{1}{2}}(\tau^{**}/\tau^*)^{\frac{1}{2}}$$
$$\times \exp\{-\frac{r}{2}(z_F-\gamma^* a-b)^T(A^*+aa^T/\tau^*)^{-1}(z_F-\gamma^* a-b)\}$$

$$\times \frac{(\beta^*)^{\alpha+\frac{N}{2}}}{\Gamma\left(\alpha+\frac{N}{2}\right)} r^{\alpha+\frac{N}{2}-1} e^{-\beta^* r} dr$$

$$= \frac{\Gamma\left(\frac{2\alpha+N+n}{2}\right)|A^*|^{-\frac{1}{2}}(\tau^{**}/\tau^*)^{\frac{1}{2}}\left(\alpha+\frac{N+n}{2}\right)^{n/2}(\beta^*)^{\alpha+N/2}}{\Gamma\left(\frac{2\alpha+N}{2}\right)(\pi(2\alpha+N+n))^{n/2}[\beta^*+(z-\gamma^* a-b)^T(A^*+aa^T/\tau^*)(z_F-\gamma^* a-b)]^{-\frac{2\alpha+N+n}{2}}} .$$

References

Anderson, T. W. (1970). *The Statistical Analysis of Time Series*, Wiley, New York.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis Forecasting and Control* (Revised edition), Holden-Day, San Francisco.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*, McGraw-Hill, New York.

McLeod, I. (1975). Derivation of the Theoretical Autocovariance Function of Autoregressive-Moving Average Time Series, *Applied Statistics 24*, 2, pp. 255-256. Correction: *26*, p. 194.

Monahan, J. F. (1977). Integration Over a Parameter Space Arising in Time Series Models, Report No. 748, Applied Math. Dept., Brookhaven Natl. Lab., Upton, New York.

Theil, H. (1971). *Principles of Econometrics*, Wiley, New York.