Discussion Paper 64

Institute for Empirical Macroeconomics
Federal Reserve Bank of Minneapolis
250 Marquette Avenue
Minneapolis, Minnesota 55480-0291

May 1992

PRIORS FOR MACROECONOMIC TIMES SERIES AND THEIR APPLICATION

John Geweke

University of Minnesota
and Federal Reserve Bank of Minneapolis

ABSTRACT

This paper takes up Bayesian inference in a general trend stationary model for macroeconomic time series with independent Student-t disturbances. The model is linear in the data, but nonlinear in parameters. An informative but nonconjugate family of prior distributions for the parameters is introduced, indexed by a single parameter which can be readily elicited. The main technical contribution is the construction of posterior moments, densities, and odds ratios using a six-step Gibbs sampler. Mappings from the index parameter of the family of prior distribution to posterior moments, densities, and odds ratios are developed for several of the Nelson-Plosser time series. These mappings show that the posterior distribution is not even approximately Gaussian, and indicate the sensitivity of the posterior odds ratio in favor of difference stationarity to the choice of the prior distribution.

Keywords and phrases: Difference stationary, Gibbs sampling, Leptokurtic distribution, Student-t distribution, Trend stationary

This work was supported in part by National Science Foundation Grant SES-8908365. I wish to acknowledge expert research assistance from Zhenyu Wang and thank Charles Nelson for providing the data.

Any opinions, findings, conclusions, or recommendations expressed herein are those of the authors and not necessarily those of the National Science Foundation, the University of Minnesota, the Federal Reserve Bank of Minneapolis, or the Federal Reserve System.

1. Introduction

Beginning with the investigation of Nelson and Plosser (1982), the propositions that most macroeconomic aggregates are trend stationary, or alternatively that they are difference stationary, have captured the attention of applied and theoretical econometricians as have few other issues. These ideas have accelerated the development of the sampling theory of estimators in the presence of nonstationarity and near-nonstationarity (Dickey and Fuller, 1981; Said and Dickey, 1984; Phillips, 1987; Sims, Stock and Watson, 1990). More recently, these questions have renewed research in Bayesian inference for time series (Zellner and Tiao, 1964; Sims, 1988; DeJong and Whiteman, 1991; Phillips, 1991; Sims and Uhlig, 1991). That basic questions about methodology are being taken up in the context of a specific empirical issue testifies to the intellectual health and vigor of econometrics. Contemporaneously with these developments, there have been rapid advances in Bayesian multiple integration which can enrich time series econometrics. The objective of this paper is to show some ways in which these advances can help address the issues of trend and difference stationarity. In doing so, it builds on a number of recent contributions, including Geman and Geman (1984), Gelfand and Smith (1990), and Geweke (1991a, 1992).

This paper breaks new methodological ground in several directions. First, it takes up Bayesian inference in an improved specification of the model of Schotman and Van Dijk (1991a, 1991b, 1992) which cannot be attacked by the essentially analytical methods of those papers or Phillips (1991). Second, it employs informative and nonconjugate priors for the parameters of interest. Third, in the light of the evidence in Geweke (1992) disturbances are leptokurtic. Finally, the paper shows how to construct exact highest posterior density regions for a model that is a nontrivial variant of the standard linear specification.

This work makes two primary substantive contributions. First, it introduces a single-parameter family of informative prior distributions for the autoregressive component of the trend stationary model. The choice of this parameter is implied by the answer to the question, "At what time interval is a uniform prior density on the unit interval for the autoregressive component plausible?". As this time interval increases the prior distribution places increasing probability on a near-nonstationary configuration, and as a corollary the posterior odds ratio in favor of difference stationarity will approach the prior odds ratio, regardless of the sample. This convergence is illustrated using the data of Nelson and Plosser (1982). Second, this work presents posterior moments, posterior densities and highest posterior density regions for these data and priors that indicate near-nonstationarity. Posterior odds ratios in favor of difference stationarity are sensitive to the choice of the parameter for the prior for the autoregressive coefficient, but never fall much below the prior odds ratio and often greatly exceed it. Posterior distributions are non-Gaussian.

The paper is organized as follows. The next section introduces the model and the likelihood function. A family of informative prior distributions is developed, and the posterior density function is derived. The posterior distribution is interpreted through successive conditioning of each of several subsets of parameters on all the other subsets of parameters. In Section 3 these conditional posterior distributions are used to construct a six-step Gibbs sampler, which generates a Markov process on the parameter space. The limiting and invariant distribution of this Markov process is the joint posterior distribution of the parameters. This section of the paper also discusses the evaluation of the numerical accuracy of Gibbs sampling-based approximations of posterior moments, shows how to construct the posterior odds ratio in favor of difference stationarity as the posterior expectation of a function of interest in the trend stationary model, and develops methods for computation of exact highest posterior density regions. Section 4 reports the empirical findings for the Nelson-Plosser data set. These include an investigation of sensitivity to the specification of the prior distribution, posterior moments and densities for the parameters of interest, and posterior odds ratios in favor of difference stationarity. The last section places the results and methods of this work in the context of a broader research agenda.

2. Prior and Posterior Distributions in the Trend Stationary Model

Schotman and Van Dijk (1991a, 1991b, 1992) have used the trend stationary model

$$y_t = \gamma + \delta t + u_t, \qquad (2.0.1a)$$

$$u_t = \rho u_{t-1} + \varepsilon_t; \quad \varepsilon_t \sim \text{IIDN}(0, \sigma^2), \qquad (2.0.1b)$$

for an observed macroeconomic time series $\{y_t\}$. This model is an alternative to the specifications employed by Nelson and Plosser (1982), Phillips (1991), and others, of the form

$$y_t = \alpha + \beta t + \rho y_{t-1} + \varepsilon_t, \qquad (2.0.2)$$

or elaborations of this form with more lagged values of the dependent variable. An important attraction of the former specification relative to the latter is that δ is the mean growth rate of $\{y_t\}$ in (2.0.1), whereas $\beta/(1-\rho)$ is the growth rate of $\{y_t\}$ in (2.0.2). Standard reference priors in (2.0.1) and (2.0.2), respectively, imply very different prior distributions for growth rates and the persistence of deviations from the trend line. Some of these differences are illustrated by Schotman and Van Dijk (1991b).

An important potential shortcoming of (2.0.1) is the restriction of all serial correlation in $\{y_t\}$ to be first order autoregressive. Indeed, using conventional frequentist model selection criteria, Nelson and Plosser (1982) found evidence against this specification for most of the fourteen macroeconomic time series they studied. Following their lead and the example of Geweke (1988b, 1989b) this study introduces additional terms in $u_{t-1} - u_{t-2}, \ldots, u_{t-k} - u_{t-k-1}$ on the right hand side of (2.0.1b). In addition, based on evidence reported in Geweke (1992), disturbances are not assumed to be Gaussian.

2.1 The likelihood function

The difference stationary model used in this research is

$$y_t = \gamma + \delta t + u_t, \tag{2.1.1a}$$

$$u_t = \rho_1 u_{t-1} + \sum_{j=2}^{5} \rho_j (u_{t-j+1} - u_{t-j}) + \varepsilon_t, \quad 0 \leq \rho_1 < 1, \tag{2.1.1b}$$

$$\{\varepsilon_t\} \text{ i.i.d.}, \quad \varepsilon_t \sim t(0, \sigma^2; \nu). \tag{2.1.1c}$$

The process $\{u_t\}$ is rendered trend stationary by the restriction on ρ_1. The truncation of the $\{\rho_j\}$ after $j = 5$ is conservative, based on the fact that for annual macroeconomic time series, lags of order three or greater provide adequate allowance for secular and business cycle behavior (Geweke, 1986, 1988b). In conjunction with the prior distribution described in Section 2.2, this truncation also loses the knife-edge character it might otherwise have. The i.i.d. Student-*t* specification requires the disturbances to be leptokurtic, but for larger values of ν the distinction is inconsequential. This specification has been lightly used in applications, although it dates back at least to work in astronomy by Jeffreys (1939) who used it for mean estimation. Fraser (1976, 1979) used this distribution in a linear model, Maronna (1976) discussed maximum likelihood estimation, and Lange, Little and Taylor (1989) have applied it using the EM algorithm. Here, we exploit the equivalence of the Student-*t* distribution to the appropriate scale mixture of normals. The latter specification has been taken up in the Bayesian literature (De Finetti, 1961; Harrison and Stevens, 1976; Ramsay and Novick, 1980; West, 1984).

The usual transformation of (2.1.1) yields

$$y_t = \gamma(1 - \rho_1) + \delta(\rho_1 - \sum_{j=2}^{5} \rho_j) + \delta(1 - \rho_1)t + \rho_1 y_{t-1} + \sum_{j=2}^{5} \rho_j (y_{t-j+1} - y_{t-j}) + \varepsilon_t, \tag{2.1.2a}$$

$$\{\varepsilon_t\} \text{ i.i.d.}, \quad \varepsilon_t \sim t(0, \sigma^2; \nu). \tag{2.1.2b}$$

Conditional on the presample values $(y_{-4}, \ldots, y_0)$, the likelihood function may be expressed,

$$\sigma^{-T} \prod_{t=1}^{T} \{(1 + \varepsilon_t^2)/\nu\sigma^2\}^{-(\nu+1)/2}, \tag{2.1.3a}$$

$$\varepsilon_t \equiv y_t - \gamma(1 - \rho_1) - \delta(\rho_1 - \sum_{j=2}^{5} \rho_j) - \delta(1 - \rho_1)t - \rho_1 y_{t-1} - \sum_{j=2}^{5} \rho_j (y_{t-j+1} - y_{t-j}.) \tag{2.1.3b}$$

There are nine parameters, which subsequently will be sorted naturally into five groups: γ, δ; ρ_1; ρ_2, ρ_3, ρ_4, ρ_5; σ; and ν.

2.2 Prior distributions

The parameter of paramount interest, on which recent Bayesian studies of trend and difference stationarity have concentrated, is ρ_1. To develop a useful family of priors consider a simplified version of (2.1.1b),

$$u_t = \rho_1 u_{t-1} + \varepsilon_t, \quad 0 \le \rho_1 < 1, \quad \varepsilon_t \sim \text{IIDN}(0, \sigma^2), \quad \sigma \text{ known.} \tag{2.2.1}$$

If $\{u_t\}$ pertains to a point in time (rather than an average over a time interval) then (2.2.1) implies

$$u_t = \alpha u_{t-r} + \varepsilon_t^{(r)}, \quad \varepsilon_t^{(r)} \sim N(0, \sigma^2(1 - \rho^{2r})/(1-\rho^2)], \quad \alpha = \rho_1^r, \tag{2.2.2}$$
$$\text{cov}(\varepsilon_t^{(r)}, \varepsilon_{t-nr}^{(r)}) = 0 \text{ for any non-zero integer n.}$$

A uniform prior distribution on [0, 1) for α in (2.2.2) implies a prior distribution for ρ_1 with density $r\rho_1^{r-1}$ on [0,1) in (2.2.1). Similarly, were the prior distribution on the autoregressive parameter uniform on [0, 1) for a time interval $\tau = n^{-1}$, n integer, the implied prior distribution for ρ_1 in (2.2.1) would have density $\tau\rho_1^{\tau-1}$ on the unit interval. The notion of a "flat prior" for ρ_1 is meaningless without reference to a sampling interval for the time series. These considerations motivate the family of priors,

$$\pi_{\rho_1}(\rho_1) = (s+1)\rho_1^s I_{[0,1)}(\rho_1); \tag{2.2.3}$$

$s = 0$ corresponds to a flat prior on the autoregressive parameter for annual data, $s = 29$ for data recorded every 30 years, $s = -11/12$ for monthly data, etc. The temporal aggregation argument is only motivating: if taken literally in (2.1.2) one would have to deal with the presence of the $y_{t-j+1} - y_{t-j}$, the nonnormality of the disturbances, and interaction between prior distributions for the other parameters and ρ_1, any one of which presents technical challenges. The empirical work here is carried through to completion using several different values of s.

If the time interval between measurements is many periods, a uniform distribution for the autoregressive parameter assigns high probability to strong persistence from one period to the next. As $s \to \infty$, the effect of the prior distribution becomes the same as a reformulation of (2.1.1) with $\rho_1 = 1$. As a corollary, the posterior odds ratio in favor of $\rho_1 = 1$ must approach the prior odds ratio, as $s \to \infty$; and a posterior odds ratio for (2.1.1) with $s = s^*$ in favor of $s = s^* + q$, s^* fixed, must approach the posterior odds ratio in favor of $\rho_1 = 1$, as $q \to \infty$. The operational ramifications of these facts will be seen in Section 4.

The trend coefficient δ displays no such sensitivity to time aggregation. The prior specification employed here is

$$\delta \sim N(\bar{\delta}, \sigma_\delta^2). \tag{2.2.4}$$

The same prior distributions are used for all macroeconomic time series studied. Since the data are in logarithms, δ indicates mean growth rate. The empirical work is carried out with $\bar{\delta} = 0$ and $\sigma_\delta = .05$. Some checks for sensitivity are reported in Section 4.1.

The prior specification for $\rho_2, \ldots, \rho_5$ is

$$\rho_j \sim N(0, \pi_0\pi_1^{j-1}), \quad \pi_0 > 0, \quad 0 < \pi_1 \leq 1. \tag{2.2.5}$$

This reflects the belief that these coefficients are not likely to be large in magnitude, and that they are smaller the greater the lag. A similar specification was employed by Doan, Litterman and Sims (1984) for vector autoregressions. In the empirical work $\pi_0 = .731$ and $\pi_1 = .342$: this implies a standard deviation of .5 for ρ_2, and .1 for ρ_5. Checks for sensitivity are reported in Section 4.1.

For the intercept γ of (2.1.1a) consider prior distributions of the form

$$\gamma \mid (\rho_1, \ldots \rho_5, \sigma, y_0) \sim N[m(y_0), V(\sigma^2, \rho_1)],$$

where $V(\sigma^2, \rho_1)$ has the property $\lim_{\rho_1 \to 1} (1-\rho_1)^2 V(\sigma^2, \rho_1) = 0$. Conditional on $(\rho_1, \ldots, \rho_5, \sigma, y_0)$ the prior distribution for the intercept in the reduced form (2.1.2) is

$$\gamma(1 - \rho_1) + \delta(\rho_1 - \sum_{j=2}^{5}\rho_j)$$

$$\sim N[(1-\rho_1)m(y_0) + \bar{\delta}(\rho_1 - \sum_{j=2}^{5}\rho_j), \; (1-\rho_1)^2 V(\sigma^2, \rho_1) + \sigma_\delta^2(\rho_1 - \sum_{j=2}^{5}\rho_j)^2]$$

and for the trend term it is

$$\delta(1 - \rho_1) \sim N[(1-\rho_1)\bar{\delta}, \; (1-\rho_1)^2\sigma_\delta^2].$$

The limiting distributions, as $\rho_1 \to 1$, are $N[\bar{\delta}(1-\sum_{j=2}^{5}\rho_j), \; \sigma_\delta^2(1-\sum_{j=2}^{5}\rho_j)^2]$ for the intercept, and $\delta = 0$ for the trend. This implies the limiting model

$$y_t = \delta(1-\sum_{j=2}^{5}\rho_j) + y_{t-1} + \sum_{j=2}^{5}\rho_j(y_{t-j+1} - y_{t-j}) + \varepsilon_t,$$

or equivalently

$$y_t = \delta + y_{t-1} + u_t, \tag{2.2.6a}$$

$$u_t = \sum_{j=1}^{4}\rho_{j+1}u_{t-j} + \varepsilon_t, \tag{2.2.6b}$$

with prior distribution $\delta \sim N(\bar{\delta}, \sigma_\delta^2)$. As $\rho_1 \to 1$ the trend stationary model (2.1.1) therefore passes smoothly to the difference stationary model (2.2.6). In the empirical work reported in Section 4, the prior distribution

$$\gamma \mid (y_0, \rho_1, \sigma) \sim N[y_0, \; \sigma^2/(1 - \rho_1^2)] \tag{2.2.7}$$

of Schotman and Van Dijk (1991a, 1991b, 1992) is employed.

With regard to the dispersion of ε_t, the reference prior distribution with density

$$\pi_\sigma(\sigma) \propto \sigma^{-1} \tag{2.2.8}$$

is assumed for σ. An exponential prior distribution with density

$$\pi_\nu(\nu) = \omega\exp(-\omega\nu) \tag{2.2.9}$$

is taken for ν. In the empirical work, $\omega = .25$, implying a prior mean of 4 and median of 2.77 for ω. These values are consistent with related findings in Geweke (1992), and the exponential form of the prior density allows ample probability for very fat tails in the distribution.

2.3 Posterior distributions

The product of the kernel densities of the independent prior distributions (2.2.3), (2.2.4), (2.2.5), (2.2.7), (2.2.8), (2.2.9), and the likelihood function (2.1.3) provides the posterior density kernel

$$p(\gamma, \delta, \rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \sigma, \nu)$$

$$= \rho_1^s \exp\{\tfrac{-1}{2}[(\delta - \bar{\delta})^2/\sigma_\delta^2) + \sum_{j=1}^{4}\pi_{j+1}^2/\pi_0\pi_1^j + (\gamma - y_0)^2(1 - \rho_1^2)/\sigma^2\}\sigma^{-(n+1)}\exp(-\omega\nu\ ) \tag{2.3.1.a}$$

$$\cdot\, \sigma^{-T}\prod_{t=1}^{T}(1 + \varepsilon_t^2/\nu\sigma^2)^{-(\nu+1)/2}, \tag{2.3.1b}$$

with ε_t defined in (2.1.3b). The parameters of the prior distribution are s, $\bar{\delta}$, σ_δ, π_0, π_1, and ω. This density is not only analytically intractable, but it is also quite likely immune to attack by Monte Carlo integration with importance sampling (Kloek and Van Dijk, 1978; Geweke, 1989a). A useful first step in a workable approach, is to exploit the equivalence between the i.i.d. Student-*t* and the independent heteroscedastic normal distribution noted in Geweke (1992).

To this end, consider an alternative specification,

$$\varepsilon_t \sim \text{IDN}(0, \sigma^2 v_t) \qquad (t = 1, \dots, T), \tag{2.3.2}$$

the v_t being fixed but unknown relative variance parameters. Given ν, the v_t have independent prior distributions

$$v_t^{-1} \sim \chi^2(\nu)/\nu, \quad t = 1, \dots, T.$$

The prior density kernel for each $w_t = v_t^{-1}$ is $w_t^{(\nu-2)/2}\exp(-\nu w_t/2)$ and the prior density kernel for v_t is

$$v_t^{-(\nu+2)/2}\exp(-\nu/2v_t). \tag{2.3.3}$$

The effect of the new specification is to change (2.3.1b) to

$$\prod_{t=1}^{T} v_t^{-(\nu+3)/2}\exp[\sum_{t=1}^{T}(\sigma^{-2}\varepsilon_t^2 + \nu)/2v_t]. \tag{2.3.4}$$

Integrate this expression with respect to $v_1, \dots, v_T$ to obtain the kernel

$$\prod_{t=1}^{T}(\sigma^{-2}\varepsilon_t^2 + \nu)^{-(\nu+1)/2},$$

which is proportional to (2.3.1b).

Consequently the model and prior may be expressed,

$$y_t = \gamma(1-\rho_1) + \delta(\rho_1 - \sum_{j=2}^{5}\rho_j) + \delta(1-\rho_1)t + \rho_1 y_{t-1} + \sum_{j=2}^{5}\rho_j(y_{t-j+1} - y_{t-j}) + \varepsilon_t\,;$$

$$\varepsilon_t \sim \text{IDN}(0, \sigma^2 v_t);$$

$$\pi_{\rho_1}(\rho_1) = (s+1)\rho_1^s;$$

$$\delta \sim N(\bar{\delta}, \sigma_\delta^2);$$

$$\gamma \mid (y_0, \rho_1, \sigma) \sim N[y_0,\ \sigma^2/(1-\rho_1^2)];$$

$$\rho_j \sim \text{IDN}(0, \pi_0\pi_1^{j-1}) \ \ (j = 2, \ldots, 5);$$

$$v_t^{-1} \mid \nu \sim \text{ID}\,\chi^2(\nu)/\nu \ \ (t = 1, \ldots, T);$$

$$\pi_\sigma(\sigma) \propto \sigma^{-1};$$

$$\nu \sim \exp(\omega).$$

2.4 Conditional posterior distributions

Consideration of conditional posterior distributions provides both insight into the structure of the posterior distribution, and a basis for efficient computation described in Section 3.

Conditional posterior distribution of γ *and* δ. Write

$$w_t \equiv y_t - \rho_1 y_{t-1} - \sum_{j=2}^{5}\rho_j(y_{t-j+1} - y_{t-j})$$

$$= \gamma(1-\rho_1) + \delta[\rho_1 - \sum_{j=2}^{5}\rho_j + (1-\rho_1)t] + \varepsilon_t \equiv \gamma z_{1t} + \delta z_{2t} + \varepsilon_t, \qquad (2.4.1)$$

$$\varepsilon_t \sim \text{IDN}(0, \sigma^2 v_t) \qquad (t = 1, \ldots, T),$$

$$w_{T+1} \equiv y_0 = \gamma + \varepsilon_{T+1} \equiv \gamma z_{1,T+1} + \varepsilon_{T+1},\ \ \varepsilon_{T+1} \sim N[0, \sigma^2/(1-\rho_1^2)],$$

$$w_{T+2} \equiv 0 = \bar{\delta} + \varepsilon_{T+2} \equiv \delta z_{2,T+2} + \varepsilon_{T+2},\ \ \varepsilon_{T+2} \sim N(0, \sigma_\delta^2).$$

The conditional posterior distribution for γ and δ is therefore bivariate normal, with mean and variance given by the usual generalized least squares expressions.

Conditional posterior distribution of $\rho_2, \ldots, \rho_5$. Write

$$w_t \equiv y_t - \gamma(1-\rho_1) - \delta\rho_1 - \delta[(1-\rho_1)t - \rho_1 y_{t-1}$$

$$= \sum_{j=2}^{5}\rho_j(y_{t-j+1} - y_{t-j} - \delta) \equiv \sum_{j=1}^{4}\rho_{j+1}z_{jt} + \varepsilon_t, \qquad (2.4.2)$$

$$\varepsilon_t \sim \text{IDN}(0, \sigma^2 v_t) \qquad (t = 1, \ldots, T),$$

$$w_{T+j} \equiv 0 = \rho_j + \varepsilon_{T+j} \equiv \rho_j z_{j,T+j} + \varepsilon_{T+j},\ \ \varepsilon_{T+j} \sim \text{IDN}(0, \pi_0\pi_1^{j-1}) \qquad (j = 1, \ldots, 4).$$

Again the posterior distribution is conditionally multivariate normal, with mean and variance given by the GLS expressions.

Conditional posterior distribution of ρ_1. The conditional posterior distribution results from the combination of the simple linear regression model

$$w_t \equiv y_t - \gamma + \delta \sum_{j=2}^{5} \rho_j - \delta t - \sum_{j=2}^{5} \rho_j (y_{t-j+1} - y_{t-j})$$

$$= \rho_1(y_{t-1} - \gamma + \delta - \delta t) + \varepsilon_t \equiv \rho_1 z_t + \varepsilon_t,$$

$$\varepsilon_t \sim \text{IDN}(0, \sigma^2 v_t) \qquad (t = 1, \dots, T),$$

with the prior density (2.2.3) for ρ_1. The conditional posterior distribution therefore has kernel density

$$\rho_1^s \exp[-(\rho_1 - \hat{\rho}_1)^2 / 2\lambda^2] I_{[0,1)}(\rho_1), \qquad (2.4.3)$$

where $\hat{\rho}_1 = \sum_{t=1}^{T} v_t^{-1} w_t z_t / \sum_{t=1}^{T} v_t^{-1} z_t^2$ and $\lambda^2 = \sigma^2 / \sum_{t=1}^{T} v_t^{-1} z_t^2$.

Conditional posterior distribution of ν. From (2.2.9) and (2.3.4) this distribution has kernel density

$$(\nu/2)^{T\nu/2} \Gamma(\tfrac{\nu}{2})^{-T} \exp(-\eta \nu), \qquad (2.4.4)$$

where $\eta = \frac{1}{2} \sum_{t=1}^{T} [\log(v_t) + v_t^{-1}] + \omega$.

Conditional posterior distribution of $v_1, \dots, v_T$. From (2.3.4), the conditional posterior density of $\psi \equiv (\sigma^{-2}\varepsilon_t^2 + \nu)/v_t$ is proportional to $\psi^{-(\nu-1)/2} \exp(-\psi/2)$. Hence

$$(\sigma^{-2}\varepsilon_t^2 + \nu)/v_t \sim \chi^2(\nu + 1). \qquad (2.4.5)$$

This result may be obtained heuristically by noting that in the prior distribution $\nu/v_t \sim \chi^2(\nu)$, that in the likelihood function for (2.3.2) $\sigma^{-2}\varepsilon_t^2/v_t$ enters in the form of the kernel density of the $\chi^2(1)$ distribution, and appealing to the reproductive property of the chi-square distribution.

Conditional posterior distribution of σ. Given all the other parameters, the posterior density kernel for σ is

$$\sigma^{-(T+1)} \exp(-\sum_{t=1}^{T} \varepsilon_t^2 / 2\sigma^2 v_t).$$

The kernel density of $\phi \equiv \sum_{t=1}^{T} (\varepsilon_t^2/v_t)/\sigma^2$ is $\phi^{-(T+1)/2} \exp(-\phi/2)$. Consequently

$$\sum_{t=1}^{T} (\varepsilon_t^2/v_t)/\sigma^2 \sim \chi^2(T), \qquad (2.4.6)$$

which has an obvious heuristic.

3. Computation of Posterior Moments and Densities

In this study the Gibbs sampler (Gelfand and Smith, 1990) is used to produce a sequence of drawings from the parameter space that is neither independent nor identically distributed, but converges in distribution to the posterior distribution whose kernel density is (2.3.1). Consistent with the discussion of Section 2.4, adopt the following notation and groupings of parameters:

$$\theta_1' = (\gamma, \delta); \qquad \theta_4 = \nu;$$
$$\theta_2' = (\rho_2, \rho_3, \rho_4, \rho_5); \qquad \theta_5' = (v_1, .., v_T)';$$
$$\theta_3 = \rho_1; \qquad \theta_6 = \sigma;$$
$$\theta' = (\theta_1', \theta_2', \theta_3, \theta_4, \theta_5', \theta_6)'.$$

The Gibbs sampling algorithm for the posterior distribution is easy to construct. Begin with an arbitrary initial value

$$\theta^{(0)} = (\theta_1^{(0)\prime}, \theta_2^{(0)\prime}, \theta_3^{(0)}, \theta_4^{(0)}, \theta_5^{(0)\prime}, \theta_6^{(0)})$$

$$\in \Theta_1 \times \Theta_2 \times \Theta_3 \times \Theta_4 \times \Theta_5 \times \Theta_6 = R^2 \times R^4 \times \{[0, 1)]\} \times R^+ \times R^{T+} \times R^+ = \Theta. \quad (3.0.1)$$

A convenient choice is the ordinary least squares estimate for θ_1, θ_2, and θ_3 (forcing the appropriate constraint on ρ_1 if need be), $\nu = 4$, $\theta_5 = (1, 1, \dots, 1)'$, $\theta_6 = s^2$. These initial values were used for all results reported in this paper, but any element of Θ may be chosen. Given $\theta^{(j)}$,

(i) Draw $\theta_1^{(j)} = (\gamma^{(j)}, \delta^{(j)})'$ from the bivariate normal distribution for γ and δ indicated by the regression (2.4.1).

(ii) Draw $\theta_2^{(j)} = (\rho_2^{(j)}, \rho_3^{(j)}, \rho_4^{(j)}, \rho_5^{(j)})'$, from the multivariate normal distribution for ρ_2, ρ_3, ρ_4, and ρ_5 indicated by the regression (2.4.2);

(iii) Draw ρ_1 from the distribution whose kernel density is given by (2.4.3). A computationally efficient method is described in Appendix A.

(iv) Draw ν from the distribution whose kernel density is given by (2.4.4); see Appendix A.

(v) Draw $v_1, \dots, v_T$ successively and independently according to the conditional posterior distributions (2.4.5).

(vi) Draw σ^2 using (2.4.6).

These six steps constitute a single pass of the Gibbs sampler. After each pass a function of interest $g(\theta^{(j)})$ can be computed, and after m passes $m^{-1} \sum_{j=1}^{m} g(\theta^{(j)})$ provides a numerical approximation to $E[g(\theta)]$.

This procedure is superficially similar to the EM algorithm, which has been used to maximize the likelihood function in a related but simpler situation by Lange, Little and Taylor (1989). Leonard (1975) used a similar approach to find an approximate posterior mode in a related problem. The superficial similarity stems from similar conditioning in each iteration. However, the Gibbs sampler produces the entire posterior distribution, not just the mode. This section takes up the justification for this procedure and some important technical details. Some of the discussion in this section closely parallels that in Geweke (1992), Section 4.

3.1 Numerical approximations

The essential characteristic of this procedure is the convergence in distribution of the continuous-state Markov chain described by (i) - (vi) to the posterior distribution.

Theorem 1. Let $\{\theta^{(j)}\}_{j=1}^{\infty}$ denote a sequence of passes for the Gibbs sampling algorithm. Then $\{\theta^{(j)}\}$ converges in distribution to the posterior distribution whose kernel density is given by (2.3.1).

Proof. The result follows from the decomposition (3.0.1) and the fact that each conditional density is positive at every point on the relevant Θ_j $(j = 1, \dots, 6)$. Letting p_j and P_j denote the conditional probability densities and probability measures respectively $(j = 1, \dots, 6)$,

(1) $p_j(\theta_j \mid \{\theta_i, i \neq j\}) > 0$ for all $\{\theta_i, i \neq j\}$.

(2) for any P_j-measurable set $A \in \Theta_j$, $P_j(A \mid \{\theta_i, i \neq j\})$ is absolutely continuous with respect to $\{\theta_i, i \neq j\}$.

Condition (1) implies that the continuous state space Markov chain induced by the Gibbs sampler is π-irreducible, aperiodic, and positive Harris recurrent. Let $[\Theta, \mathsf{A}, P^m(\cdot \mid \theta^{(0)})]$ denote the probability space induced at the end of pass m by the Gibbs sampler beginning from the initial condition $\theta^{(0)}$, and (Θ, A, P) the probability space corresponding to the posterior density $p(\cdot)$. From Theorem 3.8 of Nummelin (1984) or Corollary 1 of Tierney (1991), $\sup_{A \in \mathsf{A}} |P^m(A \mid \theta^{(0)}) - P(A)| \rightarrow 0$, and consequently $\{\theta^{(j)}\}$ converges in distribution to the posterior distribution.

Theorem 2. In addition to the assumptions of Theorem 1, suppose that the posterior expectation of $|g(\theta)|$ exists and is finite. Then

$$\bar{g}_m \equiv m^{-1}\sum_{j=1}^{m} g(\theta^{(j)}) \rightarrow E[g(\theta)],$$

where the convergence is almost sure.

Proof. The conditions of Theorem 2 imply that $\{\theta^{(j)}\}$ is ergodic. The result follows from Theorem 4.3.6 of Revuz (1975) or Theorem 3 of Tierney (1991).

The informative prior distributions developed in Section 2 guarantee the existence of all posterior moments of parameters of interest like ρ_1 and δ, to which Theorem 2 therefore applies. Formal Bayesian problems can always be cast in the form of determining the posterior expectation of a function of interest. A case in point important for this work is the posterior odds ratio taken up in Section 3.3.

For certain commonly reported posterior moments two variants on this procedure often yield much more accurate numerical approximations. For specificity, consider the case of θ_1. Given $\theta_2, \dots, \theta_6$, $E(\theta_1 \mid \theta_2, \dots, \theta_6) = (Z'V^{-1}Z)^{-1}Z'V^{-1}w \equiv \hat{\theta}_1$, where Z is the matrix of regressors and w the vector of dependent variables from (2.4.1), and $V = \text{diag}(v_1, \dots, v_T)$. Obviously $\hat{\theta}_1$ and θ_1 have the same posterior expectation, but the former has smaller posterior variance than the latter by the Rao-Blackwell theorem. This suggests that a more accurate approximation of $E(\theta_1)$ will be achieved by choosing as the function of interest $\hat{\theta}_1$ rather than θ_1 itself. Evaluating accuracy as described in Section 3.2, this indeed turns out to be the case. A similar procedure can be used to improve the numerical approximation of posterior variances (Geweke, 1992).

When conditional moments are not known analytically, this procedure cannot be used, but the closely related method of antithetic acceleration (Geweke, 1988a) can be applied. For illustration again consider the case of θ_1, but suppose now that $E[g(\theta_1) \mid \{\theta_2, \dots, \theta_6\}]$ is not a known function of $\{\theta_2, \dots, \theta_6\}$. Let θ_1^A be a drawing from the bivariate normal distribution for θ_1 implied by (2.4.1), define $\eta = \theta_1^A - \hat{\theta}_1$, and $\theta_1^B = \hat{\theta}_1 - \eta$. Then $E[g(\theta_1)] = E\{.5[g(\theta_1^A) + g(\theta_1^B)]\}$, but in many cases $\text{var}\{.5[g(\theta_1^A) + g(\theta_1^B)]\} << \text{var}[g(\theta_1)]$. If g were linear then the former variance would be zero. In fact g is not linear (if it were then the method described in the previous paragraph would be used) but under weak regularity conditions it approaches linearity in the relevant domain as sample size increases and the ratio $\text{var}\{.5[g(\theta_1^A) + g(\theta_1^B)]\}/\text{var}[g(\theta_1)]$ approaches zero (Geweke, 1988a).

3.2 Evaluation of numerical accuracy

A compelling advantage of Monte Carlo integration methods in general is that accuracy may be assessed through a central limit theorem (e.g. Geweke, 1989a, Theorem 2). In the case of the Gibbs sampler this strategy is complicated by the fact that the process $\{\theta^{(j)}\}$ is neither independently nor identically distributed. The limiting distribution of $m^{1/2}(\bar{g}_m - E[g(\theta)])$ is known to be normal under several sets of assumptions. Some require that Θ be bounded and therefore do not apply to our problem. Others (e.g. Nummelin, 1984, Corollary 7.3) pertain to bounded $g(\theta)$, and consequently apply to the computation of posterior probabilities but not posterior expectations of parameters. Even in these cases there are no subsidiary results supporting estimation of the variance of the limiting distribution. The strategy adopted here is to employ an estimated variance that would be appropriate if $\{\theta^{(j)}\}$ were a serially correlated but identically distributed process, and then make certain checks for internal consistency.

Under the assumption that $\{\theta^{(j)}\}$ is identically distributed and serially correlated the approximation of $E[g(\theta)]$ is equivalent to the classical problem of mean estimation in time series analysis. A full development is given in Geweke (1991a) and is only outlined here. Given that $g(\theta^{(j)})$ has finite mean and variance, $\bar{g}_m = m^{-1}\sum_{j=1}^{m} g(\theta^{(j)}) \rightarrow \bar{g} \equiv E[g(\theta)]$.

Under weak conditions (Hannan, 1970, Section 2.2) the spectral density $S(\omega)$ of $g(\theta^{(j)})$ exists; and $m^{1/2}(\bar{g}_m - \bar{g}) \Rightarrow N[0, S(0)]$ (Hannan, 1970, Theorem 4.11). If $\hat{S}_m(\omega)$ is a consistent (in m) estimator of $S(\omega)$, then the accuracy of $\bar{g}_m$ as an approximation of $\bar{g}$ may be assessed by the numerical standard error (NSE) $[m^{-1}\hat{S}_m(0)]^{1/2}$. Many consistent estimators of $S(0)$ are available. Technical details for the ones employed here are provided in Appendix C.

The posterior variance var(g) of $g(\theta)$ may be approximated consistently (in m) by $\widehat{\mathrm{var}}_m(g) \equiv m^{-1}\sum_{j=1}^{m} g^2(\theta^{(j)}) - \bar{g}_m^2$. Were it possible to make m i.i.d. Monte Carlo drawings directly from the posterior density, then the NSE associated with the mean of these draws would have been $[m^{-1}\widehat{\mathrm{var}}_m(g)]^{1/2}$. Following Geweke (1989a), define the squared ratio of this term to the actual NSE, $\widehat{\mathrm{var}}_m(g)/\hat{S}_m(0)$, to be the relative numerical efficiency (RNE) of the approximation $\bar{g}_m$. It indicates the relative number of drawings required to obtain a given NSE, and is a routine side computation. Some information about RNE for the moments and posterior odds ratios computed in this study is given in Appendix C.

Especially in the absence of a central limit theorem that pertains to all $g(\theta)$, and of a demonstrated consistent estimator of the variance term in the limiting normal distribution, it is important to assess the adequacy of the computed NSE's. In the work reported here that is done by repeating the computations with different initial conditions and a different seed for the random number generator. It was always the case that differences in computed posterior moments were consistent with computed NSE's and the assumption of normality.

3.3 Posterior odds ratios

Much of the recent empirical literature concerning unit roots has addressed hypotheses about ρ_1 or its conceptual equivalent in other models. The hypotheses studied have been (a) $\rho_1 < 1$; (b) $\rho_1 > 1$; (c) $\rho_1 = 1$. For reasons well stated in Schotman and Van Dijk (1991b), (c) is much more compelling than (b). Here we focus on (a) and (b); the prior distributions adapted in Section 2.2 further restrict $\rho_1 \geq 0$ in the case of (a), but trivial variants on these procedures would easily cope with prior distributions for ρ_1 extending to (-1, 0).

Begin by considering the general case of alternative hypotheses for the same model with likelihood function $L(\theta)$, $\theta \in \Theta$, that can be described by alternative prior distributions with densities $\pi_A(\theta)$ and $\pi_B(\theta)$. To fix ideas, prior distributions for ρ_1 in the trend stationary model, with density function (2.2.3) and different values of s, are examples. In the general case the posterior odds ratio in favor of hypothesis A is

$$\int_\Theta L(\theta)\pi_A(\theta)d\theta \,/\, \int_\Theta L(\theta)\pi_B(\theta)d\theta$$

$$= \int_\Theta[\pi_A(\theta)/\pi_B(\theta)]\, L(\theta)\pi_B(\theta)d\theta \,/\, \int_\Theta L(\theta)\pi_B(\theta)d\theta = E_B[\pi_A(\theta)/\pi_B(\theta)]. \tag{3.3.1}$$

The posterior odds ratio is the posterior expectation under hypothesis B of the function of interest $\pi_A(\theta)/\pi_B(\theta)$. The methods described in Section 3.1 may by applied to compute a numerical approximation of this function. The quality of the approximation will depend on the behavior of $\pi_A(\theta)/\pi_B(\theta)$ on that part of Θ where the mass of the posterior density under hypothesis B is concentrated, and in general will be better when the posterior variance of $\pi_A(\theta)/\pi_B(\theta)$, under hypothesis B, is smaller. (In general this variance need not even be finite, and this is a consideration in determining the roles of A and B in the computations.)

In the case of the alternative prior distributions,

$$\pi_A(\rho_1) = (t+1)\rho_1^t I_{[0,1)}(\rho_1), \quad \pi_B(\rho_1) = (s+1)\rho_1^s I_{[0,1)}(\rho_1), \quad t > s, \tag{3.3.2}$$

the function of interest pertinent to computation of the posterior odds ratio is

$$[(t+1)/(s+1)]\rho_1^{t-s}. \tag{3.3.3}$$

Since this function is bounded on the unit interval all its posterior moments exist. If the roles of t and s are reversed, these moments in general will not exist, and hence the configuration (3.3.2) is maintained in the empirical work.

In the case of the alternative distributions

$$\pi_A(\rho_1) = \varepsilon^{-1} I_{(1-\varepsilon,\,1)}(\rho_1), \quad \pi_B(\rho_1) = (s+1)\rho_1^s I_{[0,1)}(\rho_1),$$

the function of interest is $(s+1)^{-1}\rho_1^{-s}\varepsilon^{-1}I_{(1-\varepsilon,\,1)}(\rho_1)$, which for small values of ε is indistinguishable from

$$(s+1)^{-1}\varepsilon^{-1}I_{(1-\varepsilon,\,1)}(\rho_1). \tag{3.3.4}$$

The Gibbs sampler may be applied to approximate the posterior expectation of (3.3.4), but as ε decreases this method becomes increasingly inefficient since a very small fraction of the draws of ρ_1 from the distribution whose kernel density is given by (2.4.3) will occur in the interval $(1-\varepsilon, 1)$. Following the discussion of Section 3.1, it is computationally much more efficient to choose as the function of interest the conditional expectation of (3.3.4),

$$(s+1)^{-1}\varepsilon^{-1}\int_{1-\varepsilon}^{1}\exp[-(\rho_1-\hat{\rho}_1)^2/2\lambda^2]d\rho_1 \Big/ \int_0^1 \rho_1^s\exp[-(\rho_1-\hat{\rho}_1)^2/2\lambda^2]d\rho_1. \tag{3.3.5}$$

As $\varepsilon \to 0$, the posterior expectation of (3.3.4) approaches the posterior odds ratio in favor of $\rho_1 = 1$. Taking the same limit in (3.3.5), the posterior odds ratio is the posterior expectation of the function of interest

$$(s+1)^{-1}\exp[-(1-\hat{\rho}_1)^2/2\lambda^2] \Big/ \int_0^1 \rho_1^s\exp[-(\rho_1-\hat{\rho}_1)^2/2\lambda^2]d\rho_1. \tag{3.3.6}$$

in the difference stationary model. The one-dimensional integral in the denominator is evaluated quickly and accurately using a 21-point Gauss-Kronrod rule (IMSL, 1989, 569-572).

3.4 Posterior densities

For public reporting presentation of posterior densities is often desirable. Since the Gibbs sampler generates a representative sample of points from the posterior distribution, these densities may be approximated using conventional kernel density methods. But if the function of interest involves only a single θ_j from the partition of the parameter space, and the posterior density of the function of interest conditional on the remaining parameters is known analytically, then a much better approximation is possible. These conditions are satisfied here for the the parameters ρ_1, δ, and ν, which constitute functions of interest of the parameter subvectors θ_3, θ_1, and θ_4, respectively.

Suppose that $g(\theta)$ is a function of θ_j alone, and that the posterior density function for $g(\theta)$ conditional on $\{\theta_i, i \neq j\}$, $p_g(d \mid \{\theta_i, i \neq j\})$ is known. The unconditional posterior density function for $g(\theta)$ is $p_g(d) = E[p_g(d \mid \{\theta_i, i \neq j\}]$. This expectation may be approximated by the Gibbs sampler in the same way as that of any other function of interest. Moreover, in most cases smoothness of $p_g(d)$ is reflected in $p_g(d \mid \{\theta_i, i \neq j\})$, and the numerical approximation of $p_g(d)$ will therefore be appropriately smooth. Precisely this method is used to provide (very good) numerical approximations to the

univariate posterior densities of ρ_1, δ, and ν in Section 4.3. (In the cases of $\rho 1$ and ν, numerical one-dimensional integration of the kernel densities (2.4.3) and (2.4.4), respectively, is required. That for ρ_1 is available from (3.3.5), and that for ν is computed by transformation of
$(0, \infty)$ into (0,1) followed by evaluation using a 21-point Gauss-Kronrod rule (IMSL, 1989, 577-580).

If two functions of interest are function of the same θ_j alone, and if the conditional bivariate posterior density function for the functions of interest is known analytically, the same method may be applied. Furthermore the numerical approximations of the densities may be used in conjunction with the Gibbs sample itself, to compute highest posterior density regions to arbitrary accuracy, as follows. Given the Gibbs sample $\{\theta^{(r)}\}_{r=1}^{m}$, compute the corresponding approximations to the probability densities evaluated at these points, $p^{(s)} \equiv \hat{p}[\theta_j^{(s)}] \equiv m^{-1}\sum_{r=1}^{m} p_g(\theta_j^{(s)} \mid \{\theta_i^{(r)}, i \neq j\})$. Then sort the $p^{(s)}$ into ascending order, and compute the α'th quantile p_α^* in the obvious way. An approximate $100(1-\alpha)\%$ highest posterior density region consists of all d for which $m^{-1}\sum_{j=1}^{m} p_g(d \mid \{\theta_i^{(j)}, i \neq j\}) > p_\alpha^*$. This procedure produces the exact $100(1-\alpha)\%$ highest posterior density region as $m \rightarrow \infty$.

The bivariate posterior density of greatest interest here is that of (ρ_1, δ). The procedure just described may still be applied, but is complicated by the fact that the joint conditional distribution of ρ_1 and δ involves two subvectors of the parameter space, θ_1 and θ_3. The principal idea is to express the conditional density function for ρ_1 and δ, using a combination of analytical and numerical integration techniques. Sufficient statistics for these conditional distributions are recorded in each pass of the Gibbs sampler, and the bivariate density and highest posterior density regions are then constructed at the end. Technical details are provided in Appendix B.

4. Empirical Results for the Nelson-Plosser Data Set

These methods were applied to six of the time series studied by Nelson and Plosser (1982): real GNP, nominal GNP, real per capita GNP, unemployment, consumer prices, and velocity. Data were furnished by Charles Nelson, and the least squares estimates reported in Nelson and Plosser (1982, Table 5) were reproduced to all reported places. The sample period for the results here is the same as that used by Nelson and Plosser (1982), except that a few early observations could not be used because the model here involves five values of the lagged dependent variable, whereas the number of lags used by Nelson and Plosser (1982) varied but did not exceed four for any of these six series.

4.1 Sensitivity to the Prior Distribution

Examination of the sensitivity of results to the prior distribution for ρ_1 is a principal objective of this research, taken up in Sections 4.2 and 4.3. Here we report briefly on sensitivity to other parameters of the prior distribution subsequently fixed. These parameters are as follows.

For $j = 2, \ldots, 5$, the prior distribution of ρ_j is $IDN(0, \pi_0\pi_1^{j-1})$. For the empirical work with the six time series, $\pi_0 = .731$ and $\pi_1 = .342$, implying a standard deviation of .5 for ρ_2, .1 for ρ_5, and geometrically declining standard deviations in between. This is the "base case" of Table 1. We examine four alternative settings of these parameters, while keeping the other parameters of the prior distribution fixed at the values used in the empirical work. First, π_0 is increased by a factor of 4, doubling all standard deviations; second, π_0 is decreased by a factor of 4; third, π_0 is decreased to 2.5×10^{-5} so that the prior standard deviation of ρ_2 is .005, thus effectively constraining the coefficients on all $y_{t-j+1} - y_{t-j}$ to be zero; fourth, $\pi_0 = .25$ and $\pi_1 = 1.0$, so that the prior distribution for each of these coefficients is $N(0, .5^2)$. These four settings correspond to the four lines below the "base case" line in Table 1.

The prior distribution for the degrees-of-freedom parameter ν of the Student-*t* density of the disturbances is exponential with parameter ω and therefore has mean ω^{-1}. In the empirical work and "base case" $\omega = .25$. We examine two alternative settings: $\omega = .05$ (mean 20, or "thin tails") and $\omega = 1.0$ (mean 1, or "fat tails"). These two settings correspond to lines five and six below the "base case" in Table 1.

The prior distribution for the trend coefficient δ is $N(\bar{\delta}, \sigma_\delta^2)$. In the empirical work and "base case," the distribution has mean 0 and standard deviation .05. Since the data are in logarithms, this corresponds to a centered 95% prior confidence interval extending from a growth rate of -10% to one of +10%. Therefore this prior distribution is rather diffuse. We examine four alternative settings for $\bar{\delta}$ and σ_δ. First, σ_δ is increased to 1.0; second, σ_δ is decreased to 10^{-4}, effectively constraining δ to 0; third, the location of the base case is shifted by setting $\bar{\delta} = .03$ while maintaining $\sigma_\delta = .05$; and finally, the effects of imposing a growth rate equal to the observed rate of growth from the first to the last observation on the dependent variable are examined by setting $\bar{\delta} = .03118$ and $\sigma_\delta = 10^{-4}$. These four settings correspond to the last four lines in each of the two panels in Table 1.

The effects of these alternative settings on eight posterior moments were examined, employing the Gibbs sampler as described in Section 3.1. (Some technical details of implementation are provided in Appendix C.) For the posterior odds ratios and posterior expectations, the number of figures reported is at most one more than warranted by the numerical standard error computed as described in Section 3.2. The posterior odds ratio in favor of $\rho_1 = 1$ is computed as the posterior expectation of the function of interest (3.3.5). The posterior odds ratio in favor of "next s" is computed as the expected value of the function of interest (3.3.3), using $s = 0$ and $t = 9$ when $s = 0$ (top panel), and using $s = 9$ and $t = 29$ when $s = 9$ (bottom panel). Posterior means and standard deviations for the other parameters are computed in straightforward fashion. Several observations may be made about the results reported in Table 1.

First, there is a tendency for the posterior odds ratio in favor of $\rho_1 = 1$ to increase as the prior distribution for $\rho_2, \ldots, \rho_5$ becomes less informative. This is consistent with the fact that reasonable but more diffuse specifications have smaller posterior probability. However, the differences between the least and most restrictive prior distributions are not great. Changes in these priors have negligible effects on the posterior moments of ρ_1, δ, and ν relative to posterior standard deviations.

Second, as ω increases, raising the prior probability of smaller ν (fatter tails in the distribution of the disturbance), the posterior mean of ρ_1 decreases as does the posterior odds ratio in favor of a unit root. The effects are significant, although not overwhelming. This pattern may be related to the greater sensitivity of posterior means to outliers when disturbance distributions have thinner tails, and the implication of Perron (1989) that these same outliers tend to shift the posterior distribution of ρ_1 toward 1. Both the pattern and this interpretation bear further investigation.

Third, increased prior uncertainty about δ increases the posterior odds ratio in favor of difference stationarity. This is most evident in the effect of constraining δ to the empirical growth rate of real GNP in the sample. For realistic variations in the prior distribution for δ (i.e., excluding the last line of both panels) effects on posterior odds ratios and all moments appear very small. The reason for this effect will be discussed in the context of posterior densities, in Section 4.3.

Fourth, when δ is constrained to 0, the posterior odds ratio in favor of $\rho_1 = 1$ rises sharply. In this case the alternative to difference stationarity is simple stationarity, which is a less plausible alternative. Notice, however, that the odds ratio drops by a factor of about 8 from the top panel ($s = 0$) to the bottom panel ($s = 9$). As s increases without bound, the argument made in Section 2.2 once again applies: the posterior odds ratio must approach the prior odds ratio.

Fifth, the posterior distribution of δ is insensitive to all changes in the prior distribution except those that force the value of δ to a constant, the only example of which here is the case $\sigma_\delta = 10^{-4}$. Since δ is the coefficient of a fixed regressor, and since the specification (2.1.1) used here disentangles the trend coefficient from the autoregressive coefficient (as (2.0.2) does not, but (2.0.1) also does) this finding is not surprising.

Finally, changes in ω have two effects. The first is to shift the posterior distribution of ν substantially, $E(\nu)$ moving inversely with ω. The effects shown in Table 1 are quite large, and are consistent with the results in Geweke (1992) for the real GNP data indicating that the posterior odds ratios for pairs of fixed values of ν are within the range (.5, 2) for values of ν ranging from less than 3 to more than 20. For some other time series the evidence on ν is stronger, as will be seen. The second effect of an upward shift in ω, which assigns higher prior probability to distributions with thicker tails, is to diminish the posterior standard deviation of coefficients. This is most evident for δ in Table 1, and has been documented for a greater variety of time series and artificial data in Geweke (1992). Smaller posterior odds ratios in favor of the competing difference stationary model are consistent with these smaller posterior standard deviations.

4.2 Posterior odds ratios and moments

Using the "base case" priors for all other parameters, posterior odds ratios and moments were computed for the six indicated macroeconomic time series of Nelson and Plosser (1982). Six different prior distributions indexed by s were employed for the autoregressive coefficient ρ_1. As explained in Section 2.1 the choices for s correspond approximately to prior densities for the autoregressive coefficient that are flat on the unit interval for data recorded at various hypothetical intervals: s = -11/12, monthly; -3/4, quarterly; 0, annual; 9, every decade; 29, every 30 years; and 99, every century. Of course, the actual data used are annual in each case.

Results are reported in Table 2. For the posterior odds ratios ("P.O.R."), "Next s" refers to the value of s in the next row: e.g., in the row labeled s = - 3/4, the odds ratio is in favor of the prior specification with s = 0. Simple arithmetic shows that except for error due to numerical approximation, the "Next s" odds ratio should be the ratio of the "ρ_1 = 1" odds ratio for that row to the "ρ_1 = 1" odds ratio for the next row, a relationship that is borne out up to the number of places accuracy that numerical standard errors would indicate. These indicators of numerical accuracy are not reported here, but they are used to choose the number of digits reported in Table 2 just as they were in Table 1.

As the prior parameter s increases the posterior mean of ρ_1 increases and its posterior standard deviation decreases monotonically (within the limits of numerical accuracy) in every case. The posterior odds ratio in favor of $\rho_1 = 1$ shows a general tendency to move toward 1 as it must in the limit, but this tendency is complicated by the fact that odds ratios for high values of s are barely numerically accurate to two places in may instances. The difficulty is that the posterior distribution of (3.3.6) has substantial concentration near zero, with its expected value largely determined by high values with low posterior probability. For four of the time series there is also a tendency for the posterior odds ratio in favor of $\rho_1 = 1$ to increase above and away from 1.0 between $s = 29$ and $s = 99$. Further insight into this phenomenon will be provided by posterior densities reported and discussed in the next section.

Comparison of results for different time series in Table 2 is generally consistent with the findings of other investigators for these data: e.g., unemployment and real per capita GNP show less evidence of difference stationarity than do consumer prices and velocity. Specific comparisons provide more insight. The only published work reporting posterior probabilities or odds ratios for these data involving the hypothesis $\rho_1 = 1$, to my knowledge, is Schotman and Van Dijk (1991b). Their investigation is also one of the few to take up an informative prior distribution for ρ_1, which is essential if all of the posterior probability is not to be assigned to the difference stationary model ($\rho_1 = 1$) *a priori*. They report posterior odds ratios in favor of a unit root, when the competing hypothesis is trend stationarity with a prior distribution for ρ_1 uniform on (.8, 1.0) as follows: real GNP, .57; nominal GNP, 1.3; real per capita GNP, .53; unemployment, .20; consumer prices, 7.6; velocity, 3.1. These figures are substantially lower than those reported in Table 2 for $s = 0$ in every case, but this is due to the fact that given a prior distribution for ρ_1 uniform on the unit interval little of the posterior mass lies in (0, .8) except for the unemployment time series. Had Schotman and Van Dijk employed the unit interval prior distribution, their posterior odds ratios would increase by nearly five (less for unemployment). This brings their posterior odds ratios and those in Table 2 for $s = 0$ into rough consistency. To make the same point another way, the odds ratios reported in Table 2 for the cases $s = 9$ and $s = 29$ are in rough agreement with Schotman and Van Dijk (1991a), except for unemployment where the ratios in Table 2 all run at least three times their reported value. These remaining differences may be attributed to the richer specification of the model used here.

Posterior moments for ρ_1 and δ are consistent with those reported by Schotman and Van Dijk (1992), and with parameter estimates for more distantly related models taken up by other investigators. Posterior means and variances for ν provide new evidence, on the dispersion of the disturbances for these time series. For consumer prices, these moments strongly suggest a highly leptokurtic distribution, for which fourth moments do not exist.

Nominal GNP disturbances are almost as leptokurtic, while for the other series the posterior expectation of ν exceeds the prior mean of 4.0.

4.3 Posterior densities and highest posterior density regions

Aspects of posterior densities for the six time series are presented in Figures 1 through 18. Each figure consists of four panels. The upper right panel and the two lower panels each present a prior density (thin line) and a marginal posterior density (heavy line) for the parameter indicated on the horizontal axis. All of these densities are proper and normalized, i.e., they integrate to one. The upper left panel presents highest posterior density regions for the joint distribution of the trend coefficient, δ, and the autoregressive coefficient, ρ_1. The interior of the contour line labeled "1" has posterior probability .99, that labeled "5" has posterior probability .60, and other probabilities are listed on page 35. Scales differ from one figure to the next and axes must be examined when making comparisons, especially for the bivariate marginal densities in the upper left panels.

None of these densities is even approximately Gaussian. Marginal densities for δ are nearly symmetric, but clearly leptokurtic. As $\rho_1 \rightarrow 1$, $\mathrm{var}(\delta|\rho_1)$ increases. The source of the increase may be found in the vanishing sample variance of the term $(1 - \rho_1)t$ of the reduced form equation (2.1.2a) as $\rho_1 \rightarrow 1$. The effects many be seen in a comparison of the marginal posterior densities for δ plotted in the upper right panel: as s increases, so does dispersion in δ. (See also Table 1.) The effects are also discernible if the bivariate densities in the upper left panel are examined closely: as ρ_1 increases, so does the relative dispersion in a horizontal "slice" of the density.

An interesting aspect of the posterior distributions is the asymmetry of the bivariate density with respect to ρ_1, So too is the fact that the marginal posterior density of ρ_1 either has a local minimum near (but not at) $\rho_1 = 1$, or else it increases monotonically -- despite the existence of only a single mode of the bivariate density for ρ_1 and δ, which is interior except for a few series when $s = 29$. The source of this behavior may be found by considering the bivariate densities. Condition on δ, and let $\hat{\delta}$ denote the posterior mean of δ. As $|\delta - \hat{\delta}|$ increases, the deteriorating "fit" of the trend line $\gamma + \delta t$ increases the probability of more persistent departures from trend. This is exhibited in the upward shift of mass in the bivariate marginal density along a vertical line, as that line is moved left or right of the center of mass. This effect is also evident in Table 1, where fixing δ at its fitted value .03118 produced the smallest posterior mean for ρ_1, and fixing it at 0 produced the largest posterior mean. In many instances the distribution of δ is sufficiently disperse that there is enough mass near $\rho_1 = 1$ to produce a mode in the univariate, marginal posterior density of ρ_1 at $\rho_1 = 1$, in addition to the interior mode. Even when this is not the case the effects of the concentration near $\rho_1 = 1$ may often be seen (e.g. nominal GNP, s = 9). The local behavior in the marginal posterior density for ρ_1 near $\rho_1 = 1$ is consistent with the behavior of the posterior odds ratios in favor of difference stationarity reported in Table 2: this ratio increases going from $s = 29$ to $s = 99$ if (and only if) the marginal posterior density for ρ_1 is positively sloped at $\rho_1 = 1$ for the bivariate (ρ_1, δ) posterior densities corresponding to the prior with $s = 0$.

The marginal posterior density for ν draws a sharper contrast between series than do the posterior means and variances for this parameter reported in Table 1. This density is little affected by changes in s, for the same time series. Comparison of posterior densities for ν between series is most easily made by noting the points at which the posterior density rises above and falls below the prior density. Nominal GNP and consumer prices have the most leptokurtic disturbances: for the former $P(\nu \leq 4) \approx .75$, and for the latter $P(\nu \leq 4) \approx .91$. In all cases there is essentially no posterior mass on $\nu \in (0,1)$, and the posterior probability density at $\nu = 4$ exceeds the prior probability density. These results are consistent with strong evidence for leptokurtosis reported in Geweke (1992), where posterior odds ratios in favor of $\nu = 5$ against Gaussian disturbances range from 2.7 for velocity to 8.7×10^6 for consumer prices.

5. Conclusion

The main technical contribution of this work is to the practical application of Bayesian methods to macroeconomic time series. Beginning with a nonconjugate prior distribution in a nonlinear model with leptokurtic disturbances, it has been shown that posterior moments, odds ratios, and highest posterior densities may be computed methodically. With respect to the problem taken up, a single-parameter prior distribution for the key parameter in the trend stationary model was introduced, with the parameter implied by the answer to the question, "At what time interval is a uniform prior density on the unit interval for the autoregressive parameter plausible?". If this time interval is a year and the prior odds ratio between the hypotheses of difference and trend stationarity is 1:1, then the posterior odds ratio in favor of difference stationarity is about 2:1 for real GNP, real per capita GNP, and unemployment, 10:1 for nominal GNP, 20:1 for velocity, and over 100:1 for consumer prices. As the time interval lengthens the posterior odds ratio must necessarily converge to the prior odds ratio of 1:1, but for no time interval examined did the posterior odds ratio in favor of difference stationarity ever fall below 0.5:1. These conclusions are not very sensitive to other aspects of the prior distribution. An investigator who finds the *family* of prior distributions employed here reasonable and believes difference stationarity more plausible than trend stationarity *a priori* will find difference stationarity more plausible *a posteriori*.

Yet the results presented here are sensitive to the chosen family of prior distributions for the autoregressive coefficient. In Section 4.2 it was noted that Schotman and Van Dijk (1991b) reported posterior odds ratios in favor of difference stationarity generally lower than those given here, and the differences were traced to differences in the prior distributions. In the study of the trend and difference stationarity hypotheses there can be no such thing as a prior distribution that is uninformative, or neutral, or objective, in the sense that those words are generally used. Two factors are at work. First, there is the well known fact that the posterior odds ratio in favor of a hypothesis that restricts a parameter to a single value in the presence of an improper prior distribution for that parameter under the alternative hypothesis, must be zero (Berger and Delampady, 1987). Second, in the case of the unit roots question, the hypotheses are about the "long run," for which there is never even a single complete observation. There will always be prior distributions that dominate the data, and the macroeconomic time series record is short enough that different reasonable prior distributions may dominate the data and imply very different posterior odds ratios. Then again, they may not: there has not been enough attention to the construction of informative prior distributions that the answer to this question is clear, at least to me.

In this situation pursuit of a formulation of ignorance as absence of knowledge, in the same sense that absolute zero is absence of energy, is futile. Alternative diffuse prior distributions for the autoregressive parameter stake out positions about relative probabilities that matter for the posterior distribution. A mathematical as opposed to a subjective prespecification of ignorance is arbitrary, never neutral, and can be misleading.

An alternative and more constructive path in econometrics is the development of technology that maps informative prior distributions into posterior moments, odd ratios, and densities. Worthy goals are to make the formulation of the prior distribution as close as possible to the ways economists and knowledgeable individuals think about the problem, to allow the mapping to incorporate a rich specification of the likelihood function, and to present aspects of the posterior distribution in ways that allow the economist to understand the mapping as well as to draw substantive conclusions. I think that the computational methods illustrated in this research can be an important component of this process.

References

Berger, J.O., and M. Delampady, 1987. "Testing Precise Hypotheses," *Statistical Science* **2**: 317-352.

De Finetti, B., 1961. "The Bayesian Approach to the Rejection of Outliers," *Proceedings of the Fourth Berkeley Symposuium on Mathematical Prbability and Statistics*, vol. 1, 199-210. Berkeley: University of California Press.

DeJong, D.N., and C.H. Whiteman, 1991. "Reconsidering 'Trends and Random Walks in Macroeconomic Time Series'," *Journal of Monetary Economics*

Dickey, D.A., and W.A. Fuller, 1981. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica* **49**: 1057-1072.

Doan, T., R. Litterman and C. A. Sims, 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews* **5**(1): 1-56.

Fraser, D.A.S., 1976. "Necessary Analysis and Adaptive Inference" (with discussion), *Journal of the American Statistical Association* **71**: 99-113.

Fraser, D.A.S., 1979. *Inference and Linear Models*. New York: McGraw-Hill.

Gelfand, A.E., and A.F.M. Smith, 1990. "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association* **85**: 398-409.

Geman, S., and D. Geman, 1984. "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721-741.

Geweke, J., 1986. "The Superneutrality of Money in the United States: An Interpretation of the Evidence," *Econometrica* **54**: 1-22.

Geweke, J., 1988a. "Antithetic Acceleration of Monte Carlo Integration in Bayesian Inference," *Journal of Econometrics* **38**: 73-90.

Geweke, J., 1988b. "The Secular and Cyclical Behavior of Real GDP in Nineteen OECD Countries, 1957-1983," *Journal of Business and Economic Statistics*, **6**: 479-486.

Geweke, J., 1989a. "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica* **57**: 1317-1340.

Geweke, J., 1989b. "The Posterior Distribution of Roots in Multivariate Autoregressions," *American Statistical Association 1989 Proceedings of the Business and Economic Statistics Section.*

Geweke 1991a. "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," Federal Reserve Bank of Minneapolis Staff Report No. XX. Forthcoming in J.O. Berger, J.M. Bernardo, A.P. Dawid, and A.F.M. Smith (eds.), *Bayesian Statistics 4.* Oxford: Oxford University Press.

Geweke, J., 1991b. "Efficient Simulation from the Multivariate Normal and Student-*t* Distributions Subject to Linear Constraints," *Computer Science and Statistics: Proceedings of the Twenty-Third Sympousium on the Interface*, 571-578. Fairfax, VA: Interface Foundation of North America, Inc.

Geweke, J., 1992. "Bayesian Treatment of the Heteroscedastic Normal and Independent Student-*t* Linear Models." Federal Reserve Bank of Minneapolis working paper.

Hannan, E.J., 1970. *Multiple Time Series*. New York: Wiley.

Harrison, P.J., and C.F. Stevens, 1976. "Bayesian Forecasting" (with discussion), *Journal of the Royal Statistical Society Series* B, **38**: 205-247.

IMSL Math/Library User's Manual, 1989. Houston: IMSL, Inc.

Jeffreys, H., 1939. *Theory of Probability*. Oxford: Clarendon Press.

Kloek, T., and H.K. Van Dijk, 1978. "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo," *Econometrica* **46**: 1-19.

Lange, K.L., R.J.A. Little, and J.M.G. Taylor, "Robust Statistical Modeling Using the *t* Distribution," *Journal of the American Statistical Association* **84**: 881-896.

Leonard, T., 1975. "A Bayesian Approach to the Linear Model with Unequal Variances," *Technometrics* **17**: 95-102.

Maronna, R.A., 1976. "Robust M-Estimators of Multivariate Location and Scatter," *Annals of Statistics* **4**: 51-67.

Nelson, C.R., and C.I. Plosser, 1982. "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics* **10**: 139-162.

Nummelin, E., 1984. *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge: Cambridge University Press.

Perron, P., 1989. "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica* **57**: 1361-1402.

Phillips, P.C.B., 1987. "Time Series Regressions with a Unit Root," *Econometrica* **55**: 277-302.

Phillips, P.C.B., 1991. "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," *Journal of Applied Econometrics* **6**: 333-364.

Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, 1986. *Numerical Recipes: The Art of Scientific Computing*. Cambridge: Cambridge University Press.

Ramsay, J.O., and M.R. Novick, 1980. "PLU Robust Bayesian Decision Theory: Point Estimation," *Journal of the American Statistical Association* **75**: 901-907.

Revuz, D., 1975. *Markov Chains*. Amsterdam: North-Holland.

Said, S.E., and D.A. Dickey, 1984. "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika* **71**: 599-607.

Schotman, P.C., and H.K. Van Dijk,1991a. "A Bayesian Analysis of the Unit Root in Real Exchange Rates," *Journal of Econometrics* **49**: 195-238.

Schotman, P.C., and H.K. Van Dijk, 1991b. "On Bayesian Routes to Unit Roots," *Journal of Applied Econometrics* **6**: 387-401.

Schotman, P.C., and H. K. Van Dijk, 1992. "Posterior Analysis of Possibly Integrated Time Series with an Application to Real GNP," in P. Caines, J. Geweke, and M. Taqqu (eds.), *New Directions in Time Series Analysis, Part II*. Berlin: Springer-Verlag. (Forthcoming)

Sims, C.A., 1988. "Bayesian Scepticism on Unit Root Econometrics," *Journal of Economic Dynamics and Control* **12**: 436-474.

Sims, C.A., J.H. Stock, and M.W. Watson, 1990. "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica* **58**: 113-144.

Sims, C.A., and H. Uhlig, 1991. "Understanding Unit Rooters: A Helicopter Tour," *Econometrica* **59**

Tierney, L., 1991. "Markov Chains for Exploring Posterior Distributions." University of Minnesota School of Statistics Technical Report No. 560.

West, M., 1984. "Outlier Models and Prior Distributions in Bayesian Linear Regression," *Journal of the Royal Statistical Society* Series B **46**: 431-439.

Zellner, A., and G.C. Tiao, 1964. "Bayesian Analysis of the Regression Model with Autocorrelated Errors," *Journal of the American Statistical Association* **59**: 763-778.

Appendix A: Efficient Rejection Sampling

The general problem is to draw a sample of independent, identically distributed synthetic random variables from a target distribution with probability density function $f^*(x;\theta)$, kernel density function $f(x;\theta) = k(\theta)f^*(x;\theta)$, and support Ω. (The parameter vector θ is given.) This is to be done by drawing synthetically from a sampling distribution with probability density function $g^*(x;\alpha)$ ($\underset{p\times 1}{\alpha} \in A$, $\underset{k\times 1}{x} \in \Omega$) and kernel density function $g(x;\alpha) = h \cdot g^*(x;\alpha)$. The draw is retained with probability $c(\alpha,\theta)\, f(x;\theta) / g(x;\alpha)$, where

$$c(\alpha,\theta) \equiv \{\max_{x\in\Omega} [f(x;\theta) / g(x;\theta)]\}^{-1}.$$

This procedure is widely used in the generation of synthetic random variables; Press *et al.* (1986, 221-225) provide a brief overview. This Appendix takes up the problem of the choice of α in general and in two specific examples, given two assumptions. The first is that the family of sampling distributions indexed by α has been fixed. In fact the choice of this family is an important strategic decision, and if the performance of a rejection sampler is unsatisfactory even with α chosen as described below, one may wish to consider another family. The second assumption made here is that the objective in designing a rejection sampler is to minimize computation time which is in turn inversely proportional to the unconditional probability of retention of the draw from the sampling distribution.

The unconditional probability of retaining a draw from the sampling distribution is

$$\int_\Omega [c(\alpha,\theta)\, f(x;\theta) / g(x;\alpha]\, h^{-1}\, g(x;\alpha)dx = h^{-1} c(\alpha,\theta) \int_\Omega f(x;\theta)dx = h^{-1}k(\theta)\, c(\alpha,\theta).$$

Therefore maximization of retention probability is equivalent to maximization of $c(\alpha,\theta)$. If $c(\alpha,\theta)$ may be expressed in closed form then the problem may be solved in exactly this way. More generally, the choice of α may be expressed as the solution of the saddle point problem

$$\min_{\alpha\in A} \{\max_{x\in\Omega} [\log f(x;\theta) - \log g(x;\alpha)]\}.$$

Given the usual regularity conditions a necessary condition for solution is that α be part of a solution of the $(k+p)$-equation system

$$\partial[\log f(x;\theta) - \log g(x;\alpha)]/\partial x = 0$$
$$\partial[\log g(x;\alpha)]/\partial\alpha = 0.$$

Conditional posterior density for ρ_1. The target distribution has kernel density

$$f(x;\mu,\sigma) = x^s \exp[-(x-\mu)^2/2\sigma^2]\, I_{[0,1)}(x).$$

The sampling distribution used in this paper is $N(\nu,\sigma^2)$ truncated to the unit interval, with draws made as described in Geweke (1991b). (An obvious generalization is to use $N(\nu,\tau^2)$ in place of $N(\nu,\sigma^2)$, but the latter performs very well for the target distributions that arose, and the former leads to a much more complicated optimization problem in choosing τ^2.)

Let $g(x; \nu) \equiv \exp[-(x - \nu)^2/2\tau^2]\, I_{(0,1]}(x)$. The function

$$\log [\, f(x; \mu, \sigma^2) / g(x; \nu)] = s \log(x) - (x - \mu)^2/2\sigma^2 + (x - \nu)^2/2\sigma^2$$

$$= s \log(x) + x(\mu - \nu)/\sigma^2 + (\nu^2 - \mu^2)/2\sigma^2 \qquad (A.1)$$

is defined and concave on $(0, 1]$.

If $s\sigma^2/(\nu - \mu) \le 1$, the maximum of (A.1) is at $x^* \equiv s\sigma^2/(\nu - \mu)$. Substituting x^* for x in (A.1), the problem is to minimize

$$s \log [\, s\sigma^2 / (\nu - \mu)\,] - s + (\nu^2 - \mu^2)/2\sigma^2.$$

This function is globally convex, the first order condition is $\nu\sigma^{-2} - s(\nu - \mu)^{-1} = 0$, and the minimum is at $\nu = \frac{1}{2}[\mu + (\mu^2 + 4\sigma^2 s)^{1/2}]$. Expressing the first order condition in the form $\nu = s\sigma^2/(\nu - \mu)$, note that $x^* = \nu$. Consequently

$$\log[c(\nu; \mu, \sigma^2)] = -s \log(x^*) + (x^* - \mu)^2/2\sigma^2 - (x^* - \nu)^2/2\sigma^2$$

$$= -s \log(\nu) + (\nu - \mu)^2/2\sigma^2. \qquad (A.2)$$

Hence in this case one draws x from the distribution $N[\frac{1}{2}[\mu + (\mu^2 + 4\sigma^2 s)^{1/2}, \sigma^2]$ truncated to the unit interval, and retains the draw with probability

$$\nu^{-s} \exp[(\nu - \mu)^2/2\sigma^2] \exp\{[(x - \nu)^2 - (x - \mu)^2]/2\sigma^2\}\, x^s.$$

Finally, note that this case is identified by the condition $s\sigma^2 \le \nu - \mu$.

If $s\sigma^2 > \nu - \mu$, then (A.1) is strictly increasing on $(0, 1]$. Substituting $x^* = 1$ in (A.2), obtain $(-(1 - \mu)^2/2\sigma^2 + (1 - \nu)^2/2\sigma^2$, which is minimized by the choice $\nu = 1$, and so $\log[c(\nu; \mu, \sigma^2)] = (1 - \mu)^2/2\sigma^2$. Hence in this case one draws x from the distribution $N(1, \sigma^2)$ truncated to the unit interval, and accepts the draw with probability

$$\exp[(1 - \mu)^2/2\sigma^2] \exp\{[(x - 1)^2 - (x - \mu)^2]/2\sigma^2\}\, x^s.$$

Retention rates for some parameter combinations typical of those encountered in the research reported in this paper are shown in Table A.1 (next page).

Conditional posterior density for ν. The target distribution has kernel density

$$f(x; T, \eta) = (x/2)^{Tx/2} [\Gamma(x/2)]^{-T} \exp(-\eta x),$$

where $\eta = \frac{1}{2}\sum_{i=1}^{T} \log(v_i) + \frac{1}{2}\sum_{i=1}^{T} v_i^{-1} + \omega$. Since the function $\log(v) + v^{-1}$ is minimized at $v = 1$, $\eta \ge (T/2) + \omega$. The sampling distribution is exponential with kernel density function $g(x; \alpha) = \alpha\exp(-\alpha x)$. The function

$$Q(x, \alpha; T, \eta) \equiv \log[f(x,; T, \eta)/g(x; \alpha)]$$

$$= (Tx/2)\log(x/2) - T \log[\Gamma(x/2)] + (\alpha - \eta)x - \log(\alpha) \qquad (A.3)$$

has first derivative in x,

$$\partial Q(x, \alpha; T, \eta)/\partial x = (T/2)[\log(x/2) + 1 - \psi(x/2)] + (\alpha - \eta) \qquad (A.4)$$

where $\psi(z) = d\log[\Gamma(z)]/dz$ is the digamma, or psi, function. The function $\log(x/2) + 1 - \psi(x/2)$ is monotone decreasing from ∞ to 1 on $(0, \infty)$. Since $\eta > T/2$, (A.3) therefore has a unique, regular maximum defined by equating (A.4) to zero. This equation, together with $\partial Q(x, \alpha; T, \eta)/\partial\alpha = x - \alpha^{-1} = 0$, yields the desired solution. Substituting $\alpha = x^{-1}$ in (A.3), let x^* be the solution of

$$(T/2)[\log(x/2) + 1 - \psi(x/2)] + x^{-1} - \eta = 0,$$

which may be found by standard methods. This yields

$$\log[\, c(\alpha; T, \eta)] = -(Tx^*/2)\log(x^*/2) + T\log[\Gamma(x^*/2)] + (\eta - \alpha)x^* - \log(x^*).$$

Hence one draws from the exponential distribution with density function $x^{*-1}\exp(-x/x^*)$, and retains the draw with probability

$$[\Gamma(x^*/2)]^T (x/x^*)^{Tx/2}\, \Gamma(x/2)]^{-T} \exp[(\alpha - \eta)(x - x^*)].$$

Retention rates for some parameter combinations typical of those encountered in the research reported in this paper are shown in Table A.2.

Table A.1

Acceptance probabilities, $f(x; \mu, \sigma) = x^s \exp[-(x - \mu)^2/2\sigma^2]\, I_{[0,1)}(x)$*

	s=0	s=9	s=29	s=99
$\mu = .8,\ \sigma = .06$	1.000	.978	.948	.271
$\mu = .9,\ \sigma = .05$	1.000	.993	.968	.246
$\mu = .95,\ \sigma = .03$	1.000	.994	.990	.459

* Based on 1,000 acceptances

Table A.2

Acceptance probabilities, $f(x; T, \eta) = (x/2)^{Tx/2}\, [\Gamma(x/2)]^{-T} \exp(-\eta x)$*

$T = 60,\ \eta = 30.25$	.163	$T = 120,\ \eta = 60.25$	.111
$T = 60,\ \eta = 35.00$	.151	$T = 120,\ \eta = 70.00$	.107
$T = 60,\ \eta = 40.00$	.145	$T = 120,\ \eta = 80.00$	.113
$T = 60,\ \eta = 50.00$	.147	$T = 120,\ \eta = 100.0$	.101
$T = 60,\ \eta = 100.0$	.130	$T = 120,\ \eta = 200.0$	.105

* Based on 1,000 acceptances

Appendix B: Construction of the Bivariate Posterior Density for ρ_1 and δ

At the end of the r'th pass of the Gibbs sampler, construct

$$w_t^* = \sigma^{-1}v_t^{-1/2}[y_t - \sum_{j=2}^{5}\rho_j(y_{t-j+1} - y_{t-j}) - \gamma], \qquad z_{1t}^* = \sigma^{-1}v_t^{-1/2}(y_{t-1} - \gamma),$$

$$z_{2t}^* = \sigma^{-1}v_t^{-1/2}(t - \sum_{j=2}^{5}\rho_j), \qquad z_{3t}^* = \sigma^{-1}v_t^{-1/2}(t-1)$$

$$(t = 1, \dots, T)$$

$$w_{T+1}^* = (\sigma_\delta/\sigma)\bar{\delta}, \quad z_{2,T+1}^* = \sigma_\delta/\sigma, \quad z_{1,T+1}^* = z_{3,T+1}^* = 0.$$

The distribution of ρ_1 and δ conditional on all other parameters at the end of pass r is implicit in the regression

$$w_t^* = \rho_1 z_{1t}^* + \delta z_{2t}^* - \rho_1\delta z_{3t}^* + \varepsilon_t^*, \quad \varepsilon_t^* \sim \text{IIDN}(0, 1) \qquad (t = 1, \dots, T+1)$$

and the prior distribution for ρ_1, whose density is $(s+1)\rho_1^s I_{[0,1)}$. The kernel density of this conditional posterior distribution is

$$\tilde{p}^{(r)}(\rho_1, \delta) = \exp\{-\tfrac{1}{2}\sum_{t=1}^{T+1}(w_t^* - \rho_1 z_{1t}^* - \delta z_{2t}^* + \rho_1\delta z_{3t}^*)^2\}\rho_1^s = \exp\{-\tfrac{1}{2}Q(\rho_1, \delta)\}\rho_1^s.$$

(The superscript "(r)" on the left hand side reflects the fact that w_t^* and the z_{2t}^* are all unique to the r'th pass. We do not carry "(r)" through to the moments to avoid further cluttering the notation.) Let m_{ij} denote the raw moment $\sum_{t=1}^{T+1} z_{it}^* z_{jt}^*$, it being understood that $z_{0t}^* = w_t^*$. Then express

$$Q(\rho_1, \delta) = [\delta - \bar{\delta}(\rho_1)]^2 H(\rho_1) + A(\rho_1),$$

where

$$\bar{\delta}(\rho_1) = K(\rho_1)/H(\rho_1),$$
$$K(\rho_1) = m_{02} - \rho_1(m_{03}+m_{12}) + \rho_1^2 m_{13},$$
$$H(\rho_1) = m_{22} - 2\rho_1 m_{23} + \rho_1^2 m_{33},$$
$$A(\rho_1) = m_{00} - 2\rho_1 m_{01} + \rho_1^2 m_{11} - K(\rho_1)^2/H(\rho_1).$$

The conditional density (as opposed to the kernel) is

$$p^{(r)}(\rho_1, \delta) = \tilde{p}^{(r)}(\rho_1, \delta) / \int_0^1 \int_{-\infty}^{\infty} \tilde{p}^{(r)}(\rho_1, \delta)\, d\delta\, d\rho_1. \qquad \text{(B.1)}$$

The ten moments m_{ij} are sufficient to evaluate the numerator of (B.1) and are recorded at the end of pass r of the Gibbs sampler, along with the Gibbs sampled parameters ρ_1 and δ themselves. When all m passes of the Gibbs sampler have been completed, the numerical approximation of the bivariate density at any point (ρ_1, δ) is

$$\hat{p}(\rho_1, \delta) = m^{-1}\sum_{r=1}^{m} \tilde{p}^{(r)}(\rho_1, \delta) / d^{(r)},$$

where

$$d^{(r)} = /\int_0^1 \langle \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}[\delta - \bar{\delta}(\rho_1)]^2 H(\rho_1)\} d\delta \rangle \exp[-\tfrac{1}{2}A(\rho_1)] d\rho_1$$

$$= (2\pi)^{1/2} \int_0^1 H(\rho_1)^{-1/2} \exp[-\tfrac{1}{2}A(\rho_1)] d\rho_1.$$

The final integral is evaluated using a 21-point Gauss-Kronrod rule (IMSL, 1989, 569-572).

The bivariate density plots are prepared by evaluating $\hat{p}(\rho_1, \delta)$ on a 100 x 100 grid of points. The range is determined by evaluating $\hat{p}(\rho_1^{(r)}, \delta^{(r)})$ at all Gibbs sampled points $\{(\rho_1^{(r)}, \delta^{(r)})\}_{r=1}^{m}$, and determining the range of $\rho_1^{(r)}$ and $\delta^{(r)}$ that accommodate 99% of these points with the highest values of $\hat{p}(\rho_1^{(r)}, \delta^{(r)})$. This range is extended by 5% in each direction, except above $\rho_1 = 1$.

Appendix C: Some Computational Details

Gibbs sampling passes. Sensitivity to initial conditions in the Gibbs sampler can be reduced by discarding some of the initial passes. There is no evidence of such sensitivity in this study: even with absurd starting values, Gibbs sampled values in the next pass are typically within the 99% highest posterior density region that emerges at the conclusion of the process. Nevertheless, 200 initial passes were discarded in all cases. For the study of sensitivity to parameters of the prior distribution (Table 1), 2,000 passes were employed; for the posterior odds ratios and moments (Table 2), 10,000 passes were used; and for the posterior densities, 5,000 passes were made. In the presence of severe serial correlation every s'th pass $(s > 1)$ can be retained and the intermediate passes discarded, but there is no evidence of severe serial correlation for the posterior densities in this study. Rough ranges of relative numerical efficiencies for the parameters are as follows:

γ	.50 - 2.0	σ	.03 - .08
δ	.40 - 1.2	ν	.03 - .10
ρ_1	.15 - .70	POR, next s	.20 -. 60
$\rho_2, \dots, \rho_5$	.15 - .80	POR, $\rho_1 = 1$	.40 - 1.2

The relation between relative numerical efficiency and serial correlation is discussed in Geweke (1991a).

Computation of numerical standard errors. At the completion of all m passes, the sample $\{g(\theta^{(j)})\}_{j=1}^{m}$ is available. The spectral density at frequency zero, $S(0)$, for this series is estimated by prewhitening $\{g(\theta^{(j)})\}_{j=1}^{m}$, smoothing the periodogram around the frequency $\omega = 0$, and then recoloring.

Prewhitening is accomplished using a filter of order 10. This filter is obtained by solving the Yule-Walker equations for the autoregressive representation of order 10, using estimated autocovariances $\hat{r}(k) = m^{-1} \sum_{j=k+1}^{m} g(\theta^{(j)} - \bar{g}_m) g(\theta^{(j-k)} - \bar{g}_m)$ in lieu of the unknown true autocovariance function $r(k)$. These estimates render the 11×11 variance matrix $\hat{R} = [\hat{r}(|i-j|)]$ positive semidefinite by construction and positive definite as a practical matter. This leads to the estimated autoregressive representation

$$g(\theta^{(j)}) - \bar{g}_m = \sum_{k=1}^{10} \hat{a}_k [g(\theta^{(j-k)} - \bar{g}_m] + e_t,$$

where the $\hat{a}_k$ are obtained by solving the Yule-Walker equations using $\hat{R}$. The characteristic equation $1 - \sum_{k=1}^{10} \hat{a}_k z^k = 0$ will have no roots with modulus less than one, and as a practical matter these roots will all have modulus greater than one.

The periodogram of the prewhitened sequence $[g(\theta^{(j)}) - \bar{g}_m] - \sum_{k=1}^{10} \hat{a}_k [g(\theta^{(j-k)} - \bar{g}_m]$ is computed at the harmonic frequencies $\omega_j = 2\pi j/m$ using the fast Fourier transform, and the ordinates around $\omega = 0$ are smoothed with a Daniell window of width 20 ordinates. This yields the estimate $\hat{S}^*(0)$ of the spectral density of the prewhitened sequence. The estimate $\hat{S}(0)$ of $S(0)$ for the original sequence is $\hat{S}(0) = \hat{S}^*(0)/(1 - \sum_{k=1}^{10} \hat{a}_k)^2$. The numerical standard error for $\bar{g}_m$ is then constructed from $\hat{S}(0)$ as described in Section 3.2.

Software and hardware. Code is written in Fortran 77, making extensive use of the IMSL mathematical libraries. Execution was carried out on a Sun 4/40 IPC using 8 mb memory. Execution time is nearly proportional to the product of sample size and the total number of Gibbs sampling passes. For 10,000 passes (the number used to produce Table 2) execution time ranged from seven to ten minutes.

Table 1

Sensitivity of Some Posterior Moments to Some Parameters of the Prior Distribution

	P.O.R. in favor of		------ρ_1------		-----δx100-----		------ν------	
	$\rho_1 = 1$	Next s	Mean	s.d.	Mean	s.d.	Mean	s.d.
Real GNP, s=0								
Base case*	2.7	3.0	.851	(.076)	3.070	(0.286)	5.4	(3.1)
s.d. (ρ_2)=1.0	4.	3.4	.863	(.076)	3.053	(0.316)	5.7	(4.2)
s.d. (ρ_2)=.25	2.5	3.2	.861	(.070)	3.074	(0.310)	6.3	(3.8)
s.d. (ρ_2)=.005	2.5	3.8	.880	(.065)	3.096	(0.338)	5.8	(3.3)
s.d. (ρ_j)=.5	2.5	3.2	.856	(.076)	3.049	(0.295)	4.8	(3.7)
$\omega = .05$	3.1	3.4	.862	(.077)	3.043	(0.372)	22.	(18.)
ω=1.0	1.8	2.7	.842	(.071)	3.056	(0.246)	2.72	(1.15)
σ_δ=1.0	2.9	3.0	.850	(.075)	3.067	(0.319)	5.6	(3.3)
$\sigma_\delta=10^{-4}$	151.	9.20	.9905	(.0095)	10^{-6}	(10^{-2})	7.6	(4.5)
$\bar{\delta}$=.03, σ_δ=.05	3.6	3.27	.860	(.074)	3.062	(0.323)	5.9	(3.8)
$\bar{\delta}$=.03118, $\sigma_\delta=10^{-4}$	.5	2.3	.829	(.070)	3.1180	(.0001)	4.7	(3.1)
Real GNP, s=9								
Base case*	.89	.81	.906	(.064)	3.063	(0.422)	6.4	(3.9)
s.d.(ρ_2)=1.0	1.1	.92	.913	(.062)	3.060	(0.439)	5.6	(3.4)
s.d.(ρ_2)=.25	.68	.81	.907	(.062)	3.072	(0.416)	7.1	(4.8)
s.d.(ρ_2)=.005	1.1	.86	.915	(.055)	3.11	(0.41)	6.1	(3.3)
s.d.(ρ_j)=.5	1.2	.96	.915	(.063)	3.09	(0.425)	5.0	(3.3)
ω=.05	1.03	.90	.911	(.064)	3.018	(0.495)	21.	(23.)
ω=1.0	.8	.71	.895	(.067)	3.069	(0.370)	3.1	(1.4)
σ_δ=1.0	1.	.77	.901	(.066)	3.079	(0.434)	5.8	(3.6)
$\sigma_\delta=10^{-4}$	19.	2.550	.9912	(.0088)	10^{-4}	(10^{-2})	8.7	(5.0)
$\bar{\delta}$=.03, σ_δ=.05	.7	.74	.900	(.063)	3.07	(0.40)	5.8	(3.7)
$\bar{\delta}$=3.1180, $\sigma_\delta=10^{-4}$	.30	.50	.879	(.063)	3.1180	(.0001)	5.2	(3.0)

*Configuration of prior parameters for the base case: for ρ_j ($j = 2, \ldots, 5$) priors are independent zero-mean normal, with standard deviations declining geometrically from 0.5 for ρ_2 to 0.1 for ρ_5; for δ, prior is normal with mean $\bar{\delta} = 0$ and standard deviation $\sigma_\delta = .5$; for ν, prior is exponential with parameter $\omega = .25$ (mean $\omega^{-1} = 4.0$). The prior density for ρ_1 is $(s+1)\rho_1^s$ on the unit interval.

Table 2

Posterior Odds Ratios and Moments for Six Macroeconomic Time Series*

	P.O.R. in favor of		------ρ_1------		-----δx100-----		------ν------	
	$\rho_1 = 1$	Next s	Mean	s.d.	Mean	s.d.	Mean	s.d.
Real GNP								
s=-11/12	24.	2.917	.848	(.075)	3.057	(0.295)	5.4	(3.5)
s=-3/4	10.1	3.534	.848	(.075)	3.057	(0.297)	5.6	(3.5)
s=0	2.3	3.14	.856	(.074)	3.055	(0.297)	5.6	(3.5)
s=9	.87	.79	.904	(.064)	3.057	(0.427)	5.8	(3.7)
s=29	.99	.87	.9602	(.0368)	3.072	(0.625)	6.51	(3.70)
s=99	1.21		.9902	(.0104)	3.106	(0.842)	6.79	(4.26)
Nominal GNP								
s=-11/12	108.	2.9693	.9405	(.0370)	5.611	(0.593)	3.29	(1.94)
s=-3/4	46.	3.827	.9431	(.0368)	5.611	(0.623)	3.23	(1.74)
s=0	11.1	6.23	.9435	(.0366)	5.611	(0.644)	4.00	(1.90)
s=9	1.66	1.44	.9551	(.0324)	5.680	(0.697)	3.2	(1.9)
s=29	1.21	1.03	.9714	(.0242)	5.785	(0.804)	3.22	(1.82)
s=99	1.23		.9902	(.0098)	6.00	(1.03)	3.30	(1.87)
Real per capita GNP								
s=-11/12	17.	2.9088	.834	(.077)	1.766	(0.274)	5.66	(3.65)
s=-3/4	7.9	3.489	.834	(.077)	1.770	(0.271)	5.5	(3.7)
s=0	2.1	2.72	.839	(.077)	1.758	(0.288)	5.6	(3.5)
s=9	.68	.72	.894	(.069)	1.755	(0.412)	6.2	(3.9)
s=29	1.1	.88	.9588	(.0391)	1.758	(0.634)	6.7	(4.0)
s=99	1.22		.9901	(.0104)	1.76	(0.88)	7.4	(4.3)
Unemployment								
s=-11/12	17.	2.890	.803	(.089)	-1.14	(1.06)	5.9	(3.3)
s=-3/4	6.1	3.402	.807	(.089)	-1.14	(1.07)	6.2	(3.6)
s=0	1.8	2.19	.810	(.088)	-1.15	(1.09)	5.8	(3.5)
s=9	.67	.64	.882	(.075)	-1.25	(1.58)	6.0	(3.5)
s=29	1.18	.87	.9576	(.0413)	-1.35	(2.36)	6.0	(3.7)
s=99	1.29	1.030	.9903	(.0102)	-1.32	(2.97)	5.9	(3.8)

Table 2 (continued)

	P.O.R. in favor of $\rho_1 = 1$	Next s	------ρ_1------ Mean	s.d.	-----δx100----- Mean	s.d.	------ν------ Mean	s.d.
Consumer prices								
s=-11/12	1.67×10^3	2.99735	.9945	(.0047)	1.103	(0.631)	2.67	(1.14)
s=-3/4	5.3×10^2	3.9827	.9946	(.0045)	1.093	(0.608)	2.70	(1.15)
s=0	133.	9.530	.9949	(.0044)	1.12	(0.62)	2.59	(0.99)
s=9	14.8	2.716	.9949	(.0044)	1.119	(0.630)	2.67	(1.09)
s=29	5.24	2.44	.9949	(.0044)	1.12	(0.63)	2.78	(1.27)
s=99	2.34	1.437	.9962	(.0034)	1.17	(0.67)	2.51	(1.06)
Velocity								
s=-11/12	2.8×10^2	2.9807	.9621	(.0293)	-0.963	(0.400)	7.8	(4.5)
s=-3/4	91.	3.887	.9628	(.0287)	-0.966	(0.392)	7.2	(3.8)
s=0	21.	7.29	.9623	(.0288)	-0.970	(0.386)	7.7	(4.4)
s=9	3.0	1.78	.9688	(.0257)	-0.967	(0.418)	7.8	(4.3)
s=29	1.62	1.26	.9779	(.0199)	-0.931	(0.456)	8.0	(4.6)
s=99	1.25	1.060	.9910	(.0093)	-0.856	(0.541)	7.3	(4.1)

*For the posterior odds ratios and posterior expectations, at most the rightmost digit is uncertain because of the innacuracy of the numerical approximation, as indicated by the numerical standard error which was computed but is not reported here.

Key to Figures 1 through 18

These figures each have four panels. The upper right panel and the two lower panels each present a prior density (thin line) and a marginal posterior density (heavy line) for the parameter indicated on the horizontal axis. All of these densities are proper and normalized, i.e., they integrate to one. The upper left panel presents highest posterior density regions for the joint distribution of the trend coefficient, δ, and the autoregressive coefficient, ρ_1. The correspondence between the numbered contour lines and the probabilities for this panel is as follows.

Interior(s) of contour line(s) numbered --	has posterior probability --
1	.99
2	.95
3	.90
4	.75
5	.60
6	.40
7	.25
8	.10
9	.05
10	.01

Figure 1

Posterior Densities for Real GNP, s = 0

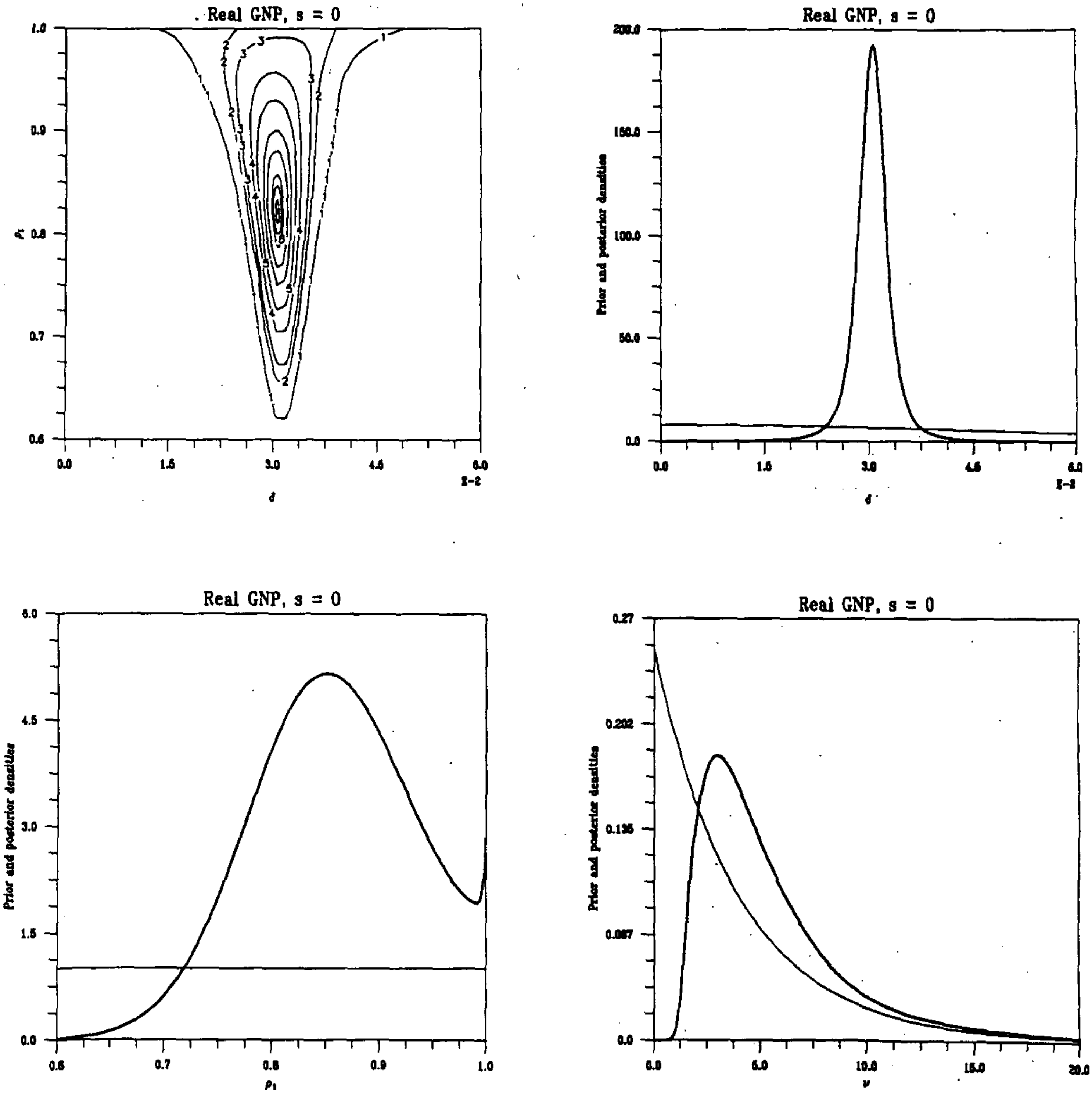

Figure 2

Posterior Densities for Real GNP, s = 9

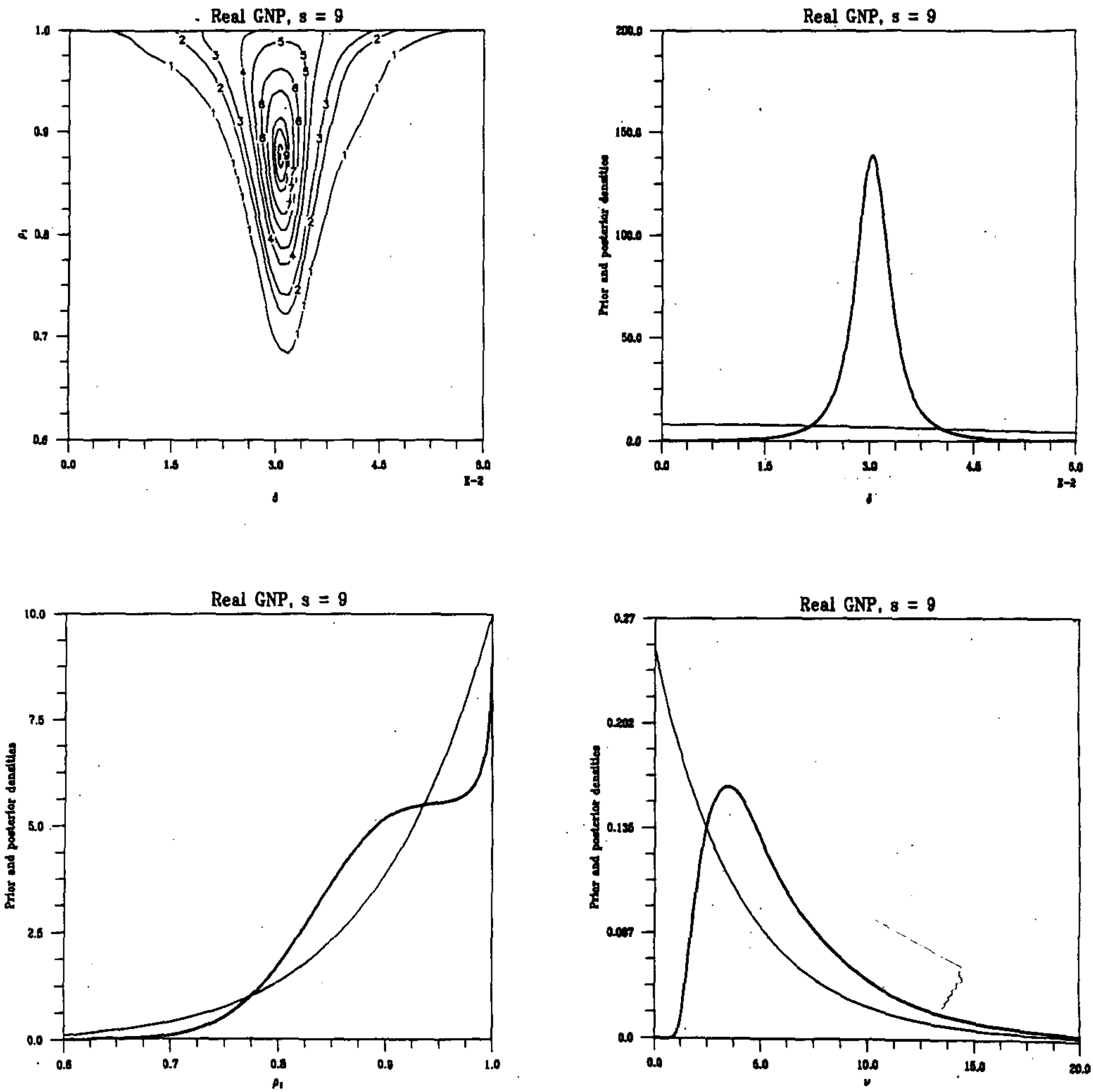

Figure 3

Posterior Densities for Real GNP, s = 29

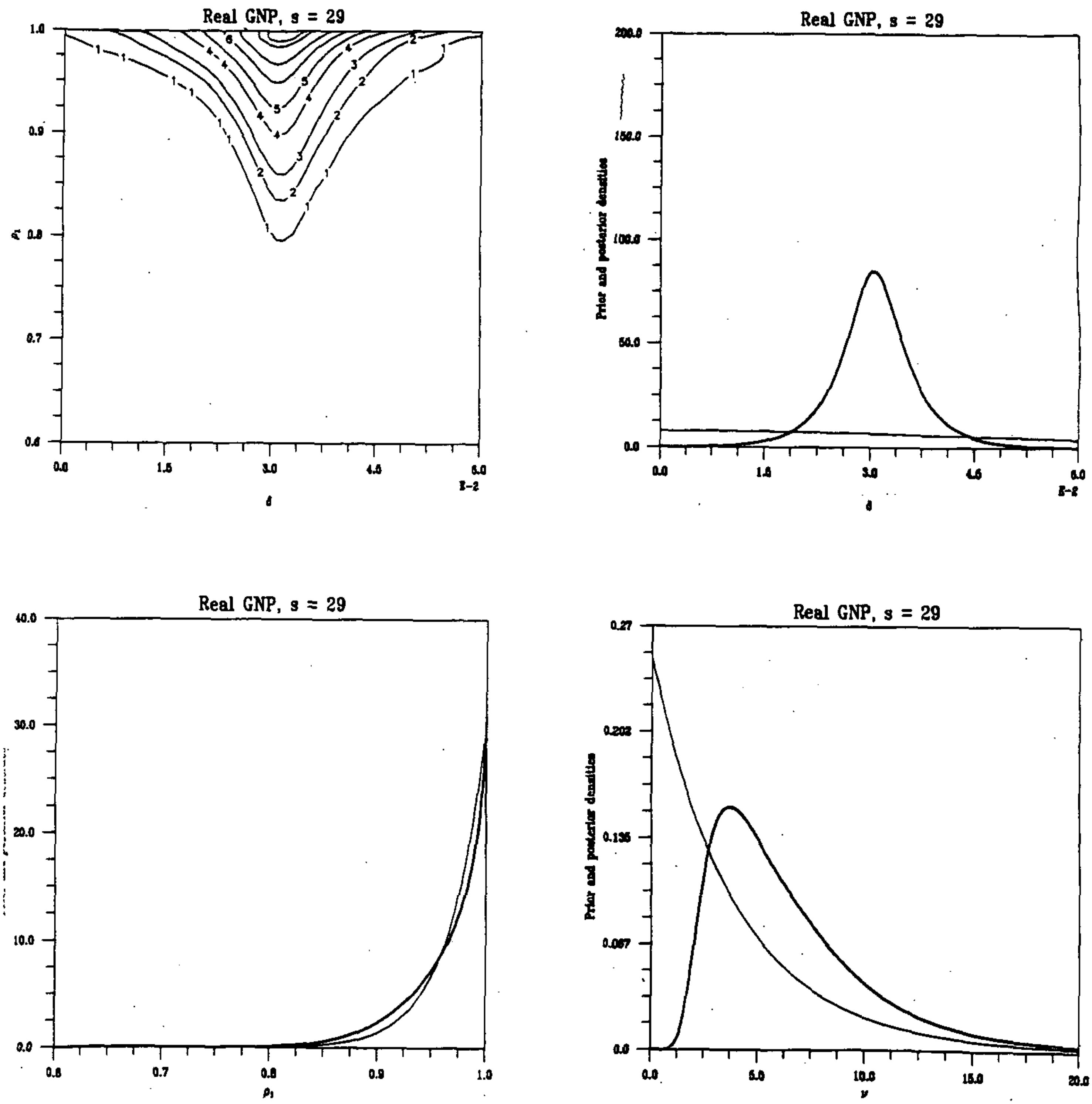

Figure 4

Posterior Densities for Nominal GNP, s = 0

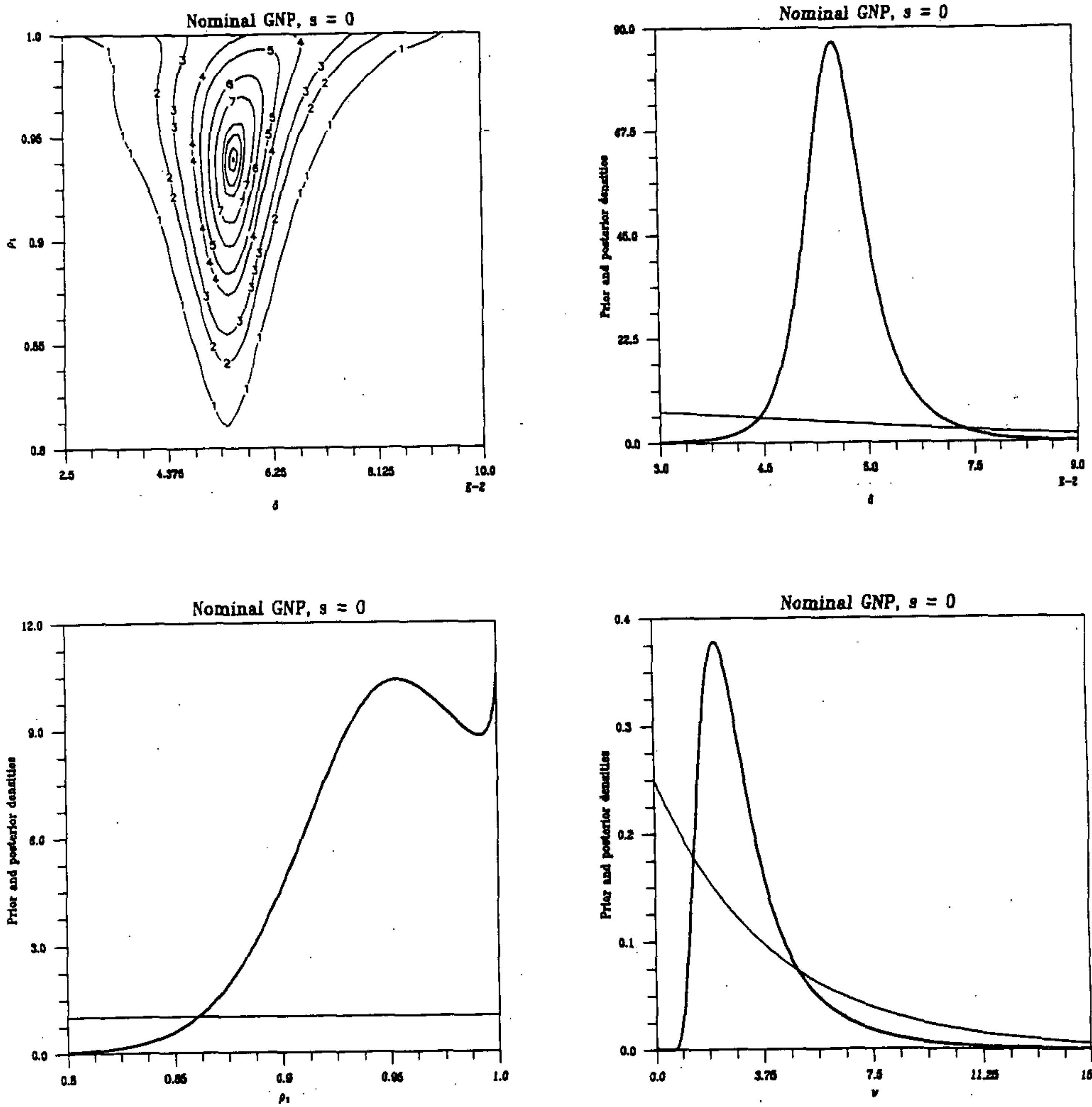

Figure 5

Posterior Densities for Nominal GNP, s = 9

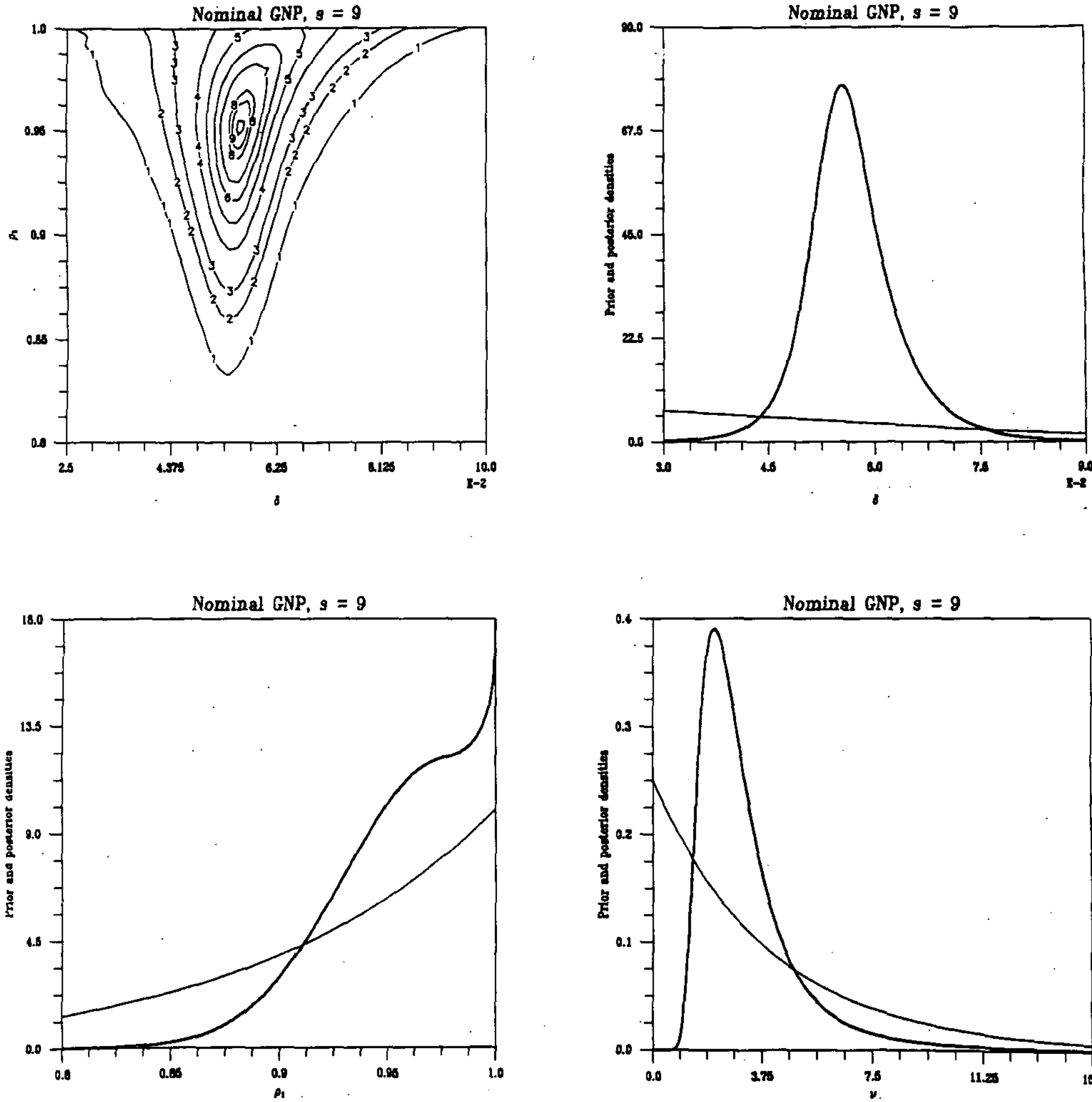

Figure 6

Posterior Densities for Nominal GNP, s = 29

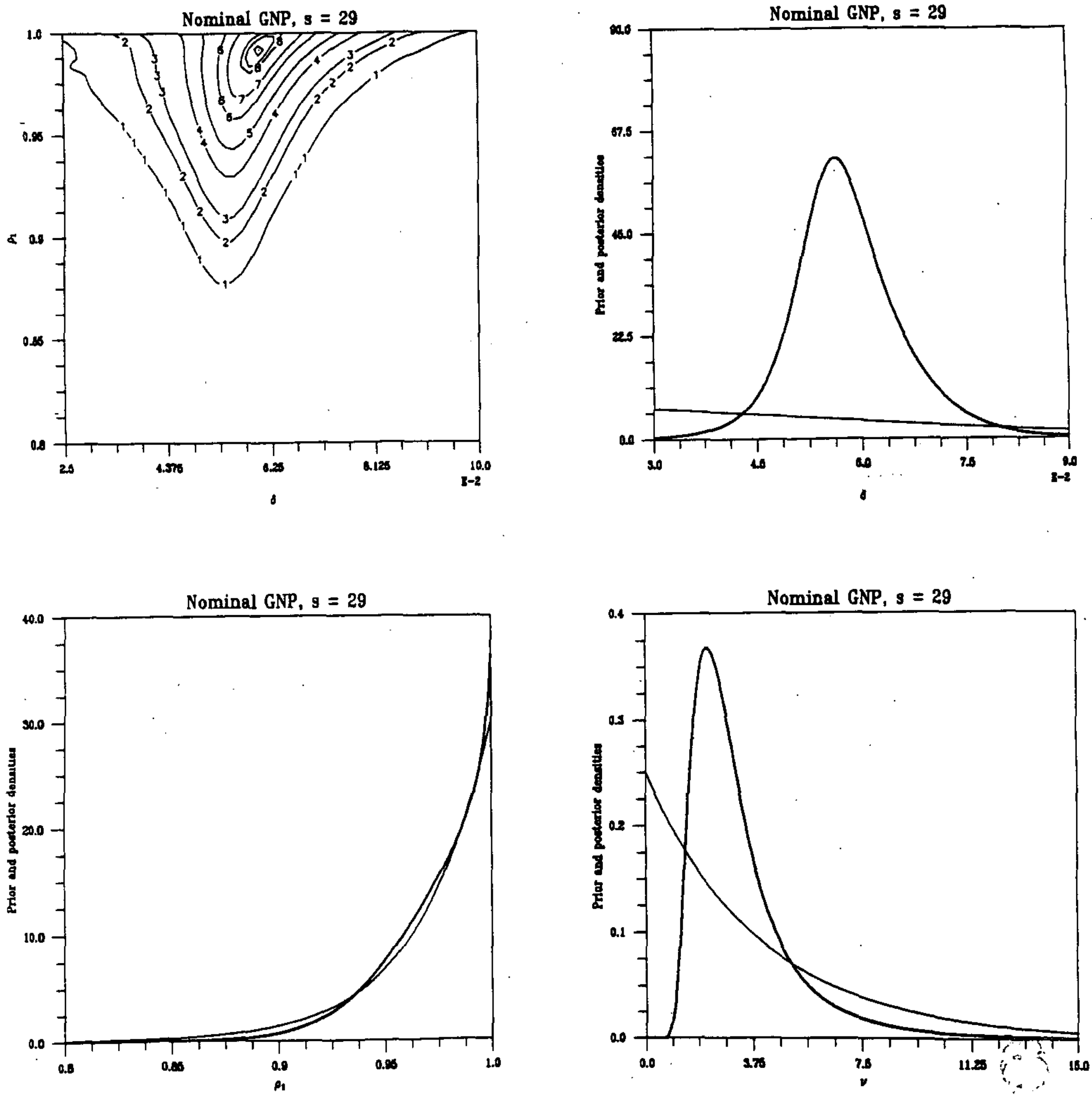

Figure 7

Posterior Densities for Real per capita GNP, s = 0

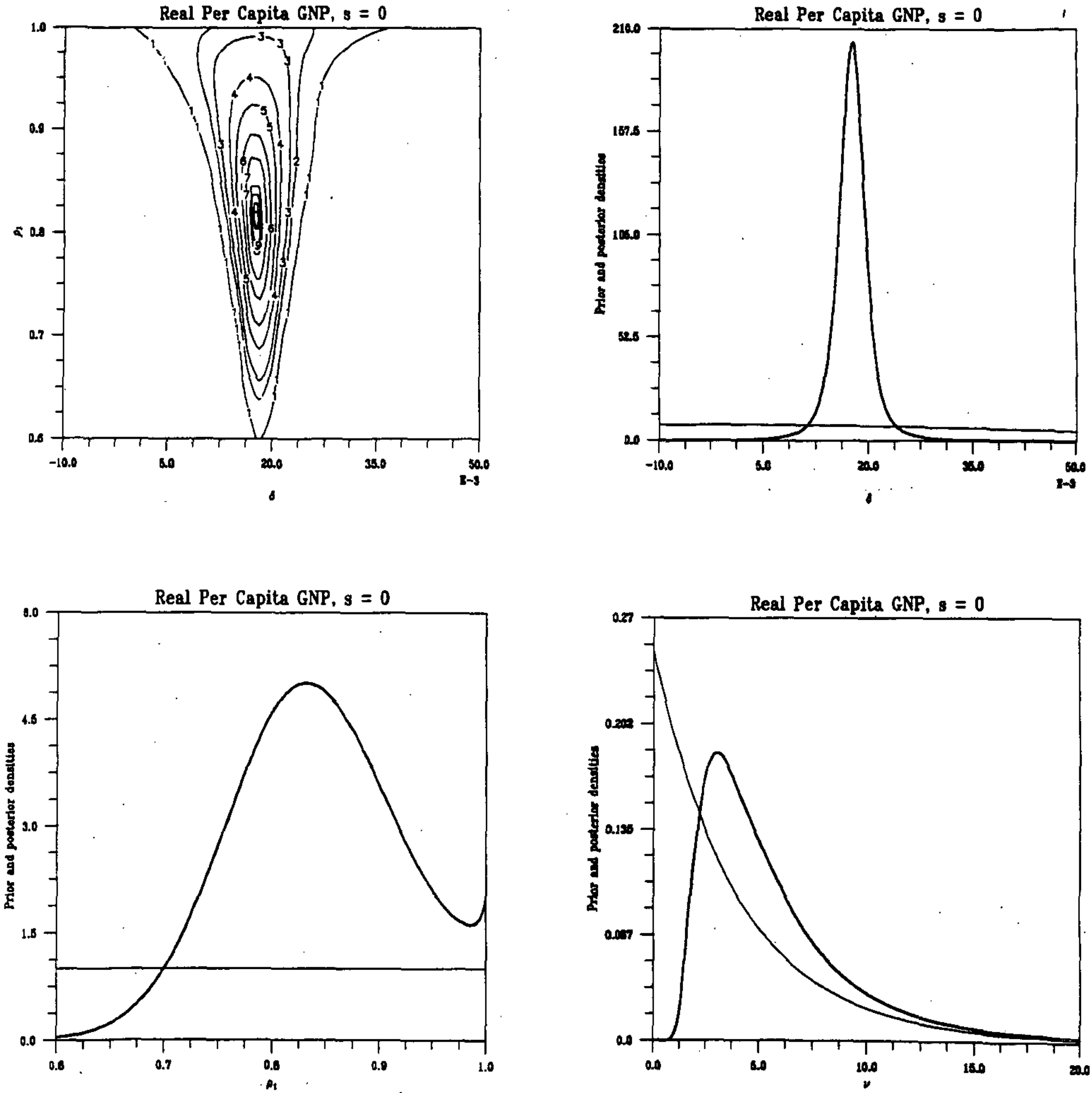

Figure 8

Posterior Densities for Real per capita GNP, s = 9

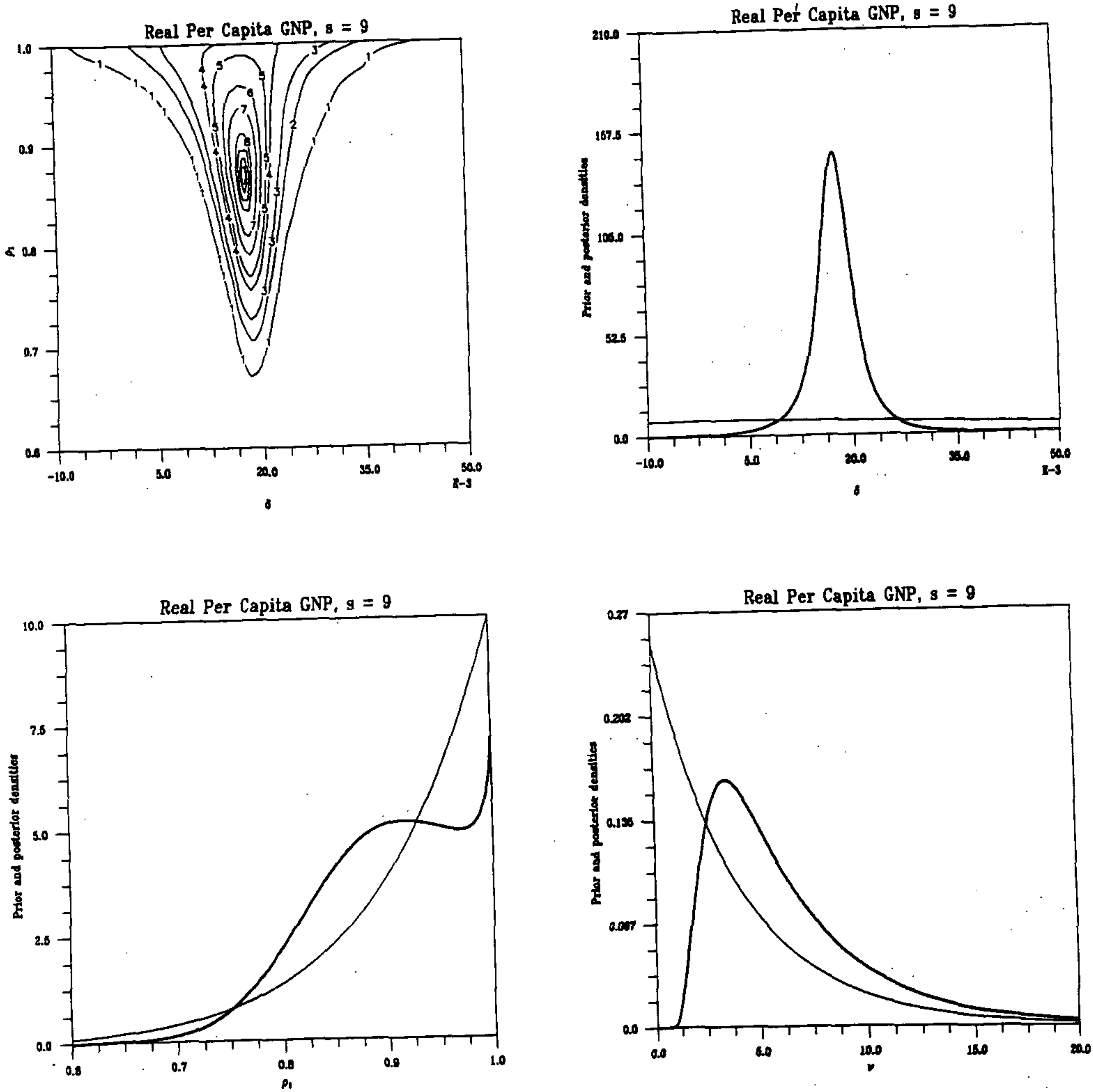

Figure 9

Posterior Densities for Real per capita GNP, s = 29

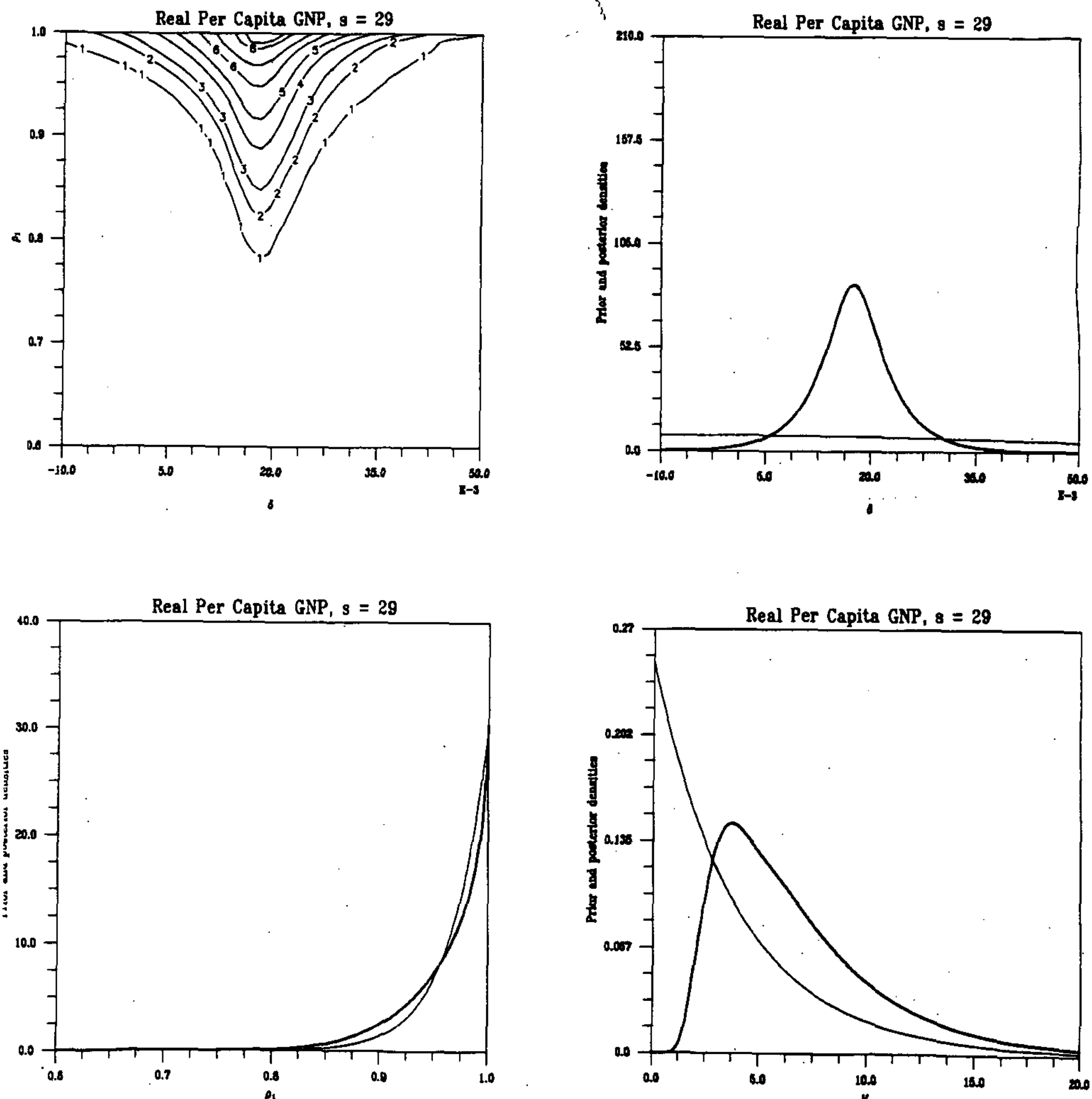

Figure 10

Posterior Densities for Unemployment, $s = 0$

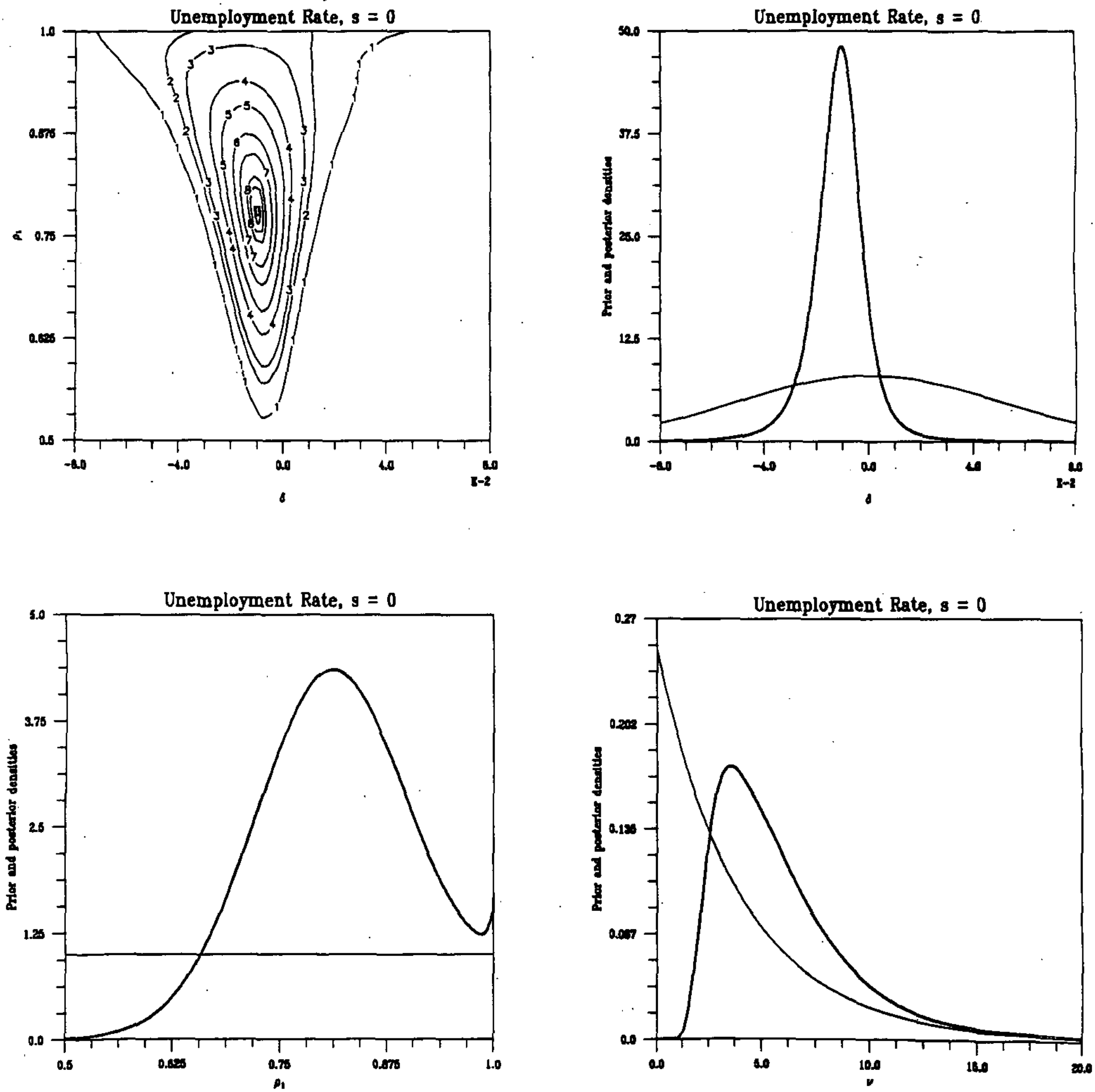

Figure 11

Posterior Densities for Unemployment, $s = 9$

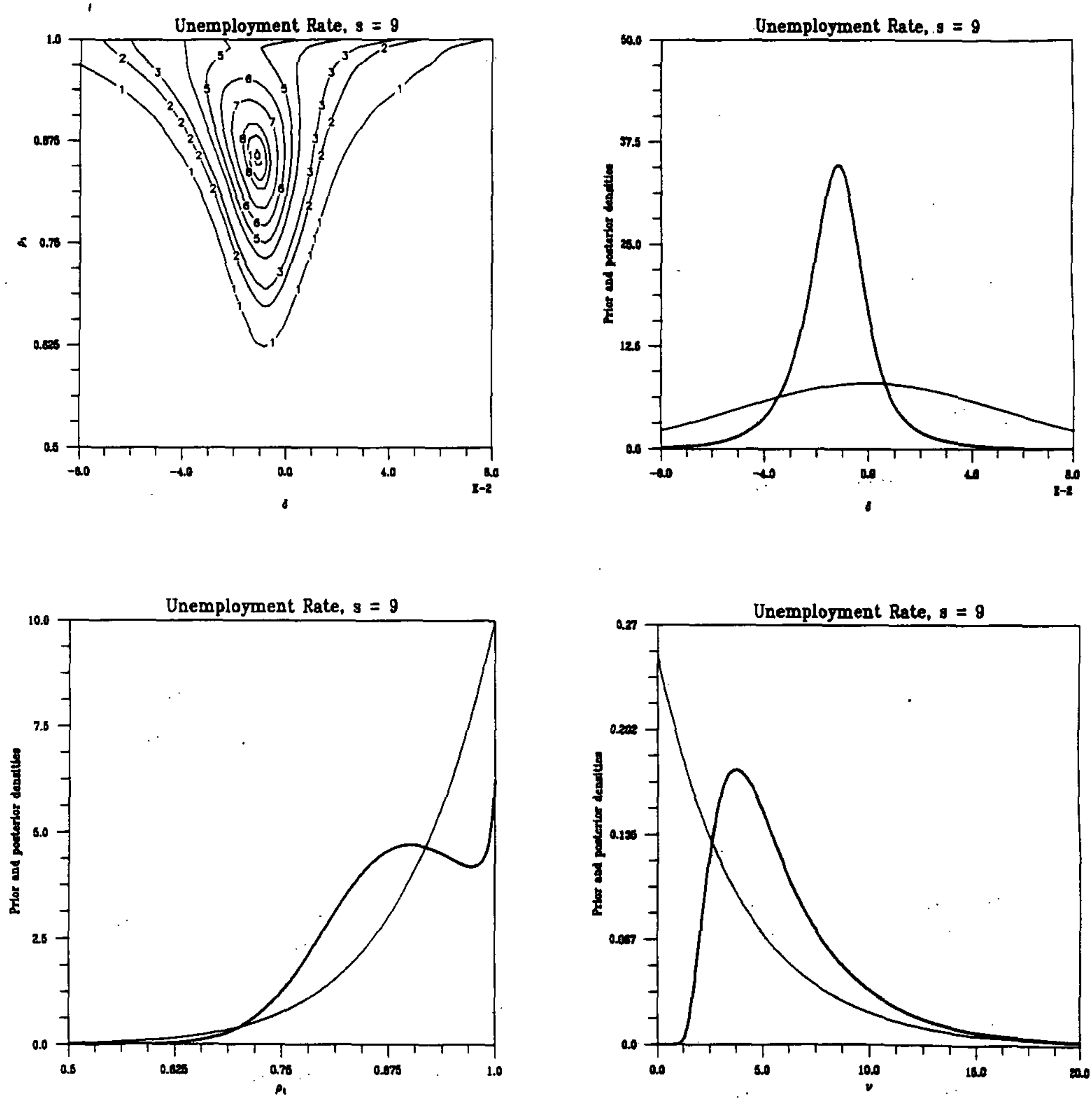

Figure 12

Posterior Densities for Unemployment, s = 29

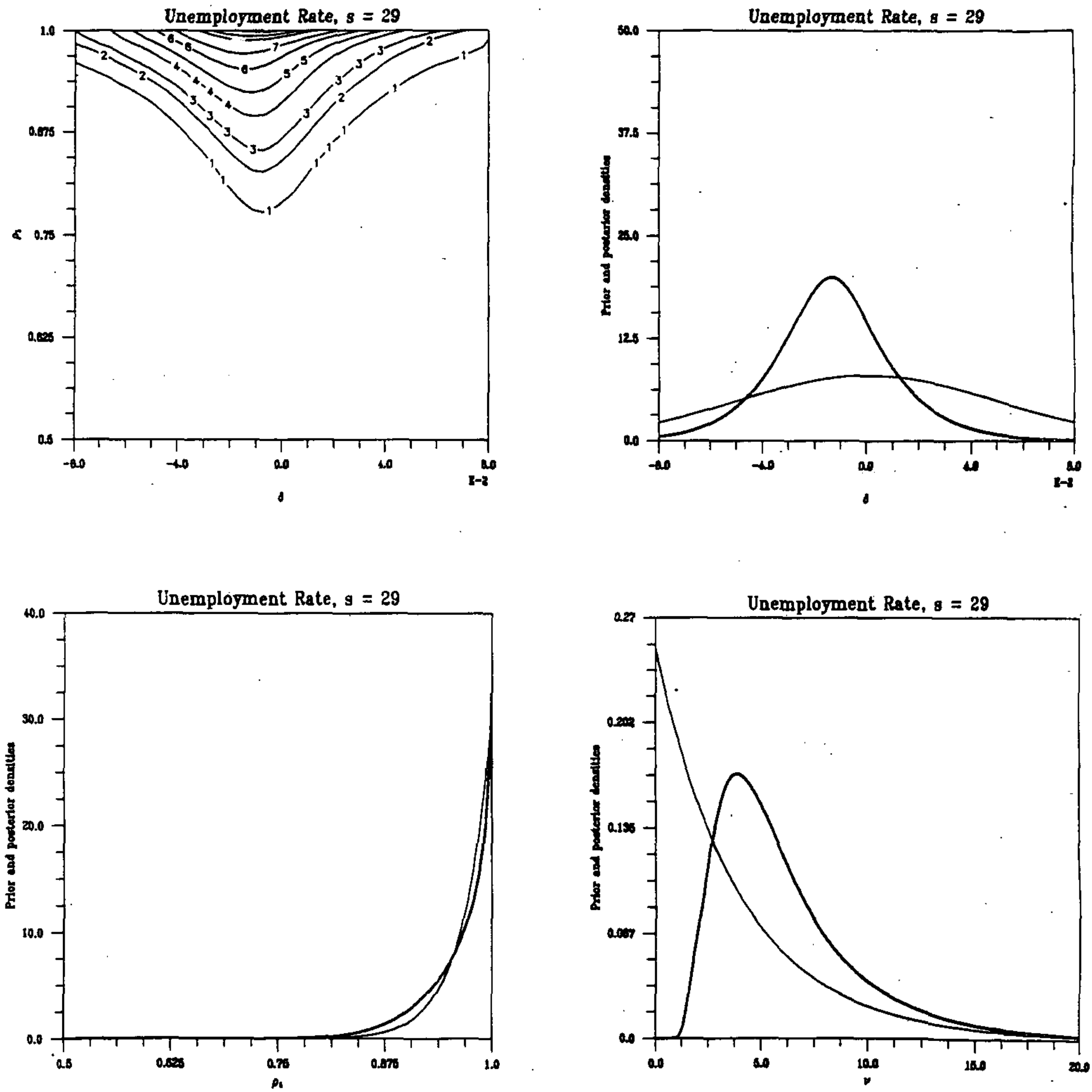

Figure 13

Posterior Densities for Consumer prices, s = 0

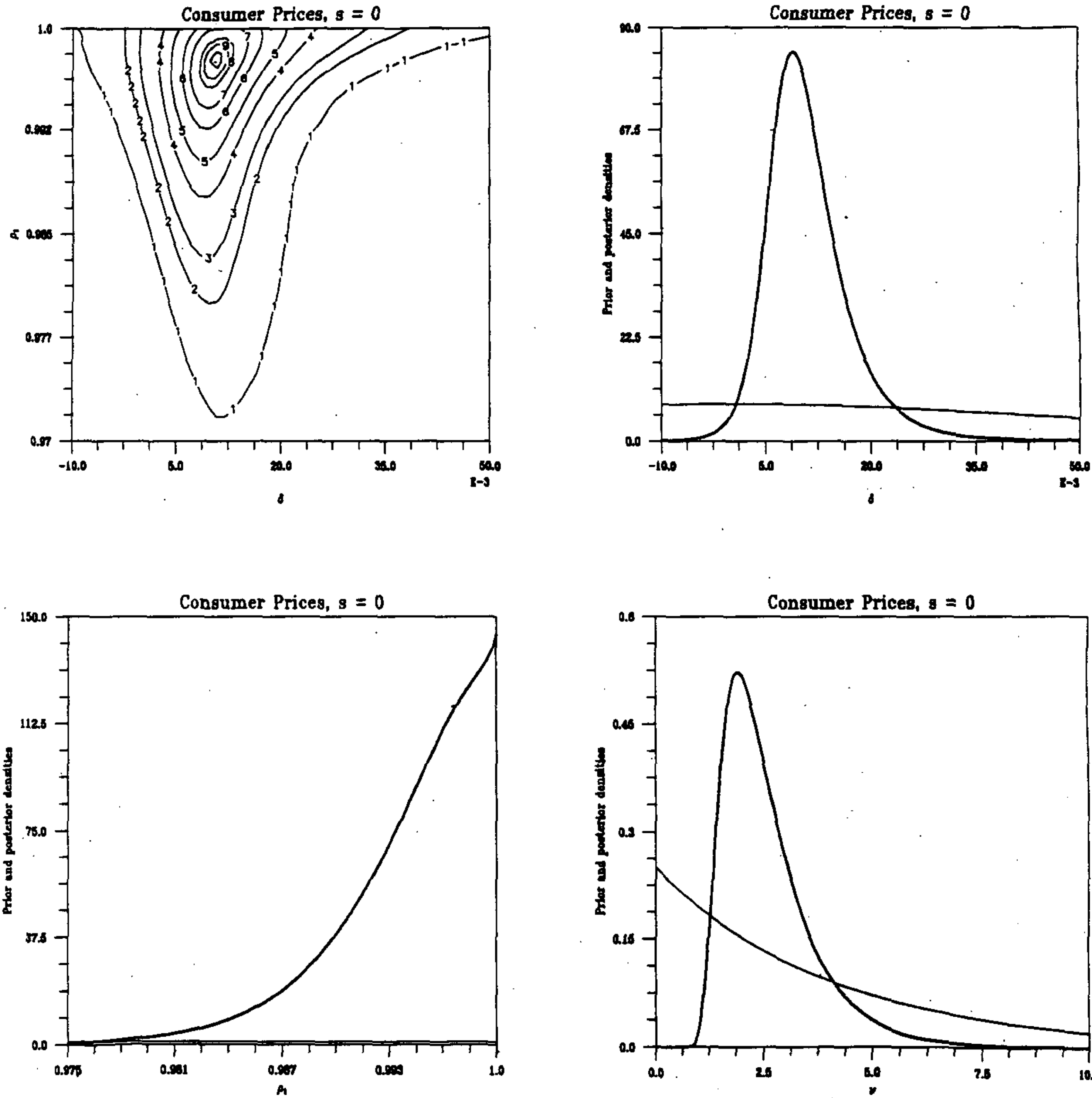

Figure 14

Posterior Densities for Consumer prices, s = 9

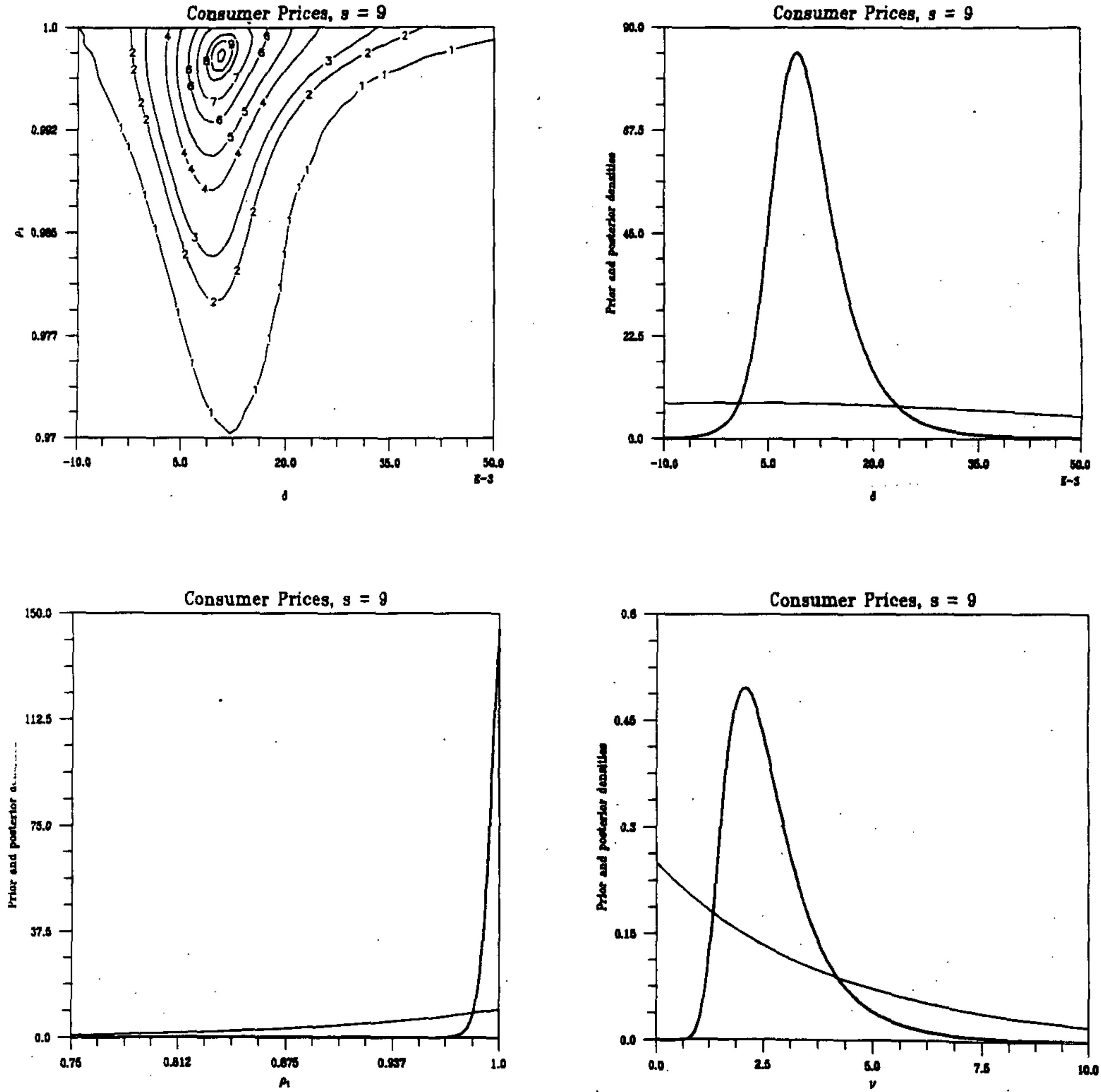

Figure 15

Posterior Densities for Consumer prices, s = 29

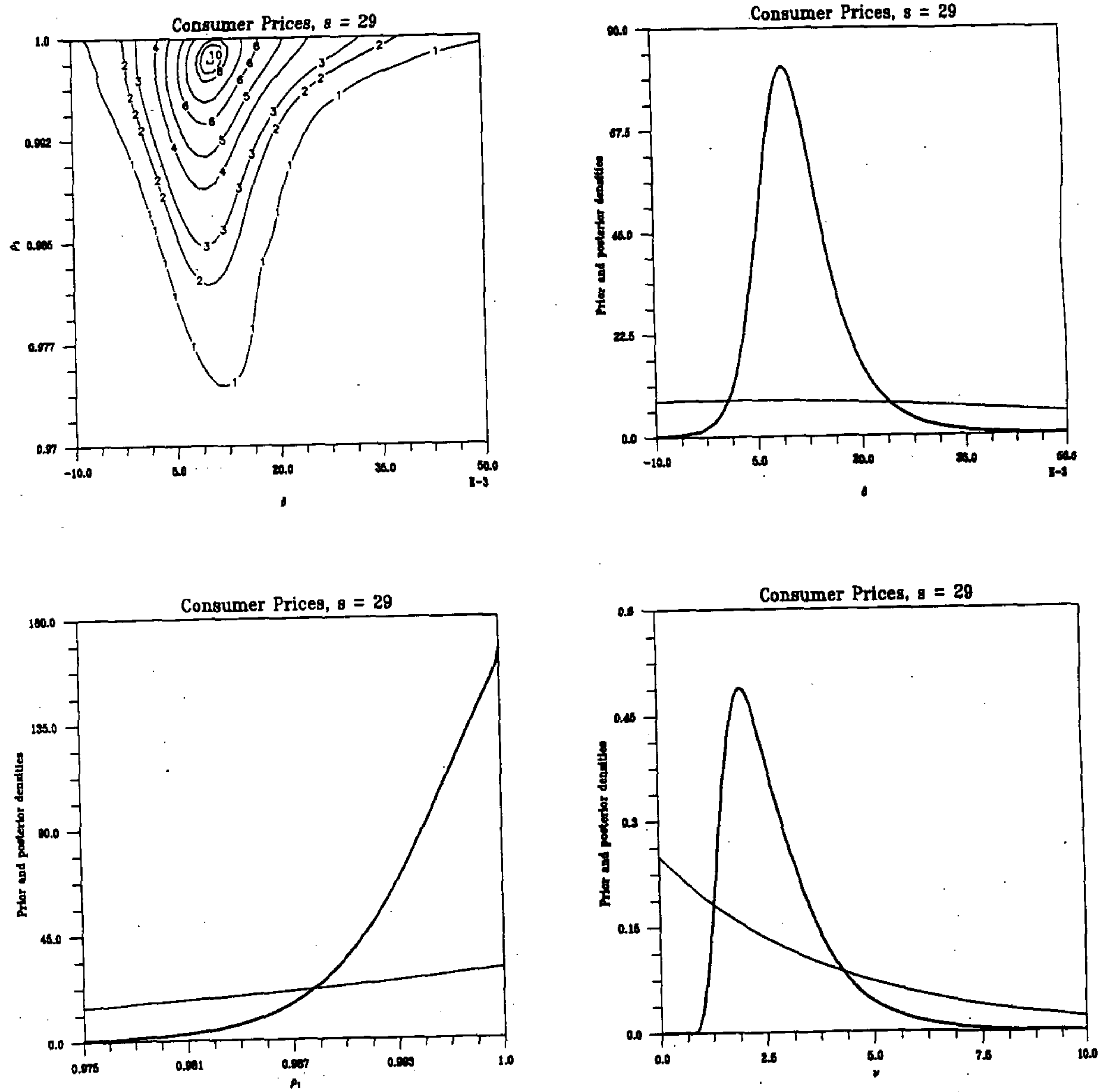

Figure 16

Posterior Densities for Velocity, $s = 0$

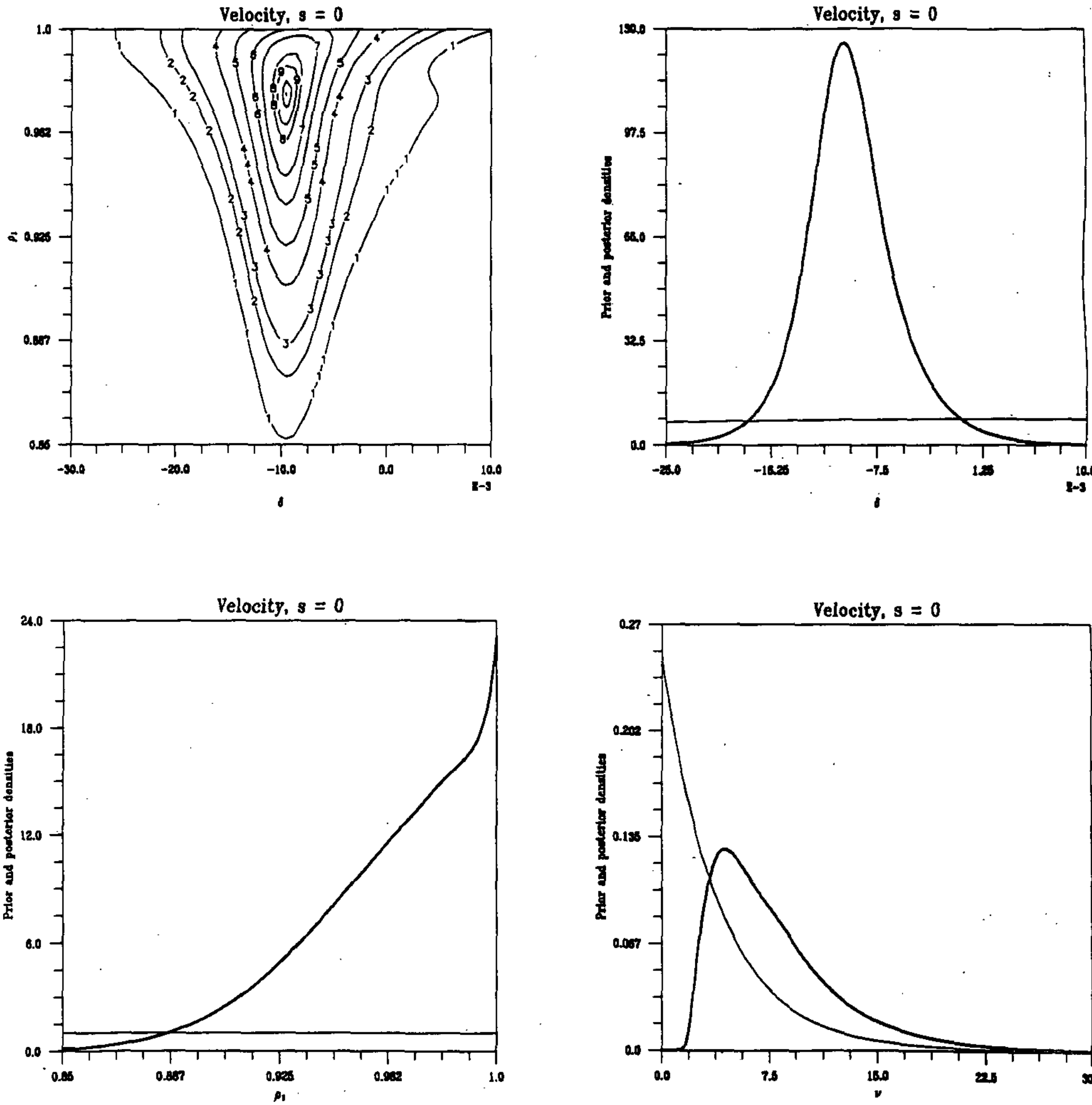

Figure 17

Posterior Densities for Velocity, s = 9