Testing for Integration using Evolving Trend and Seasonals Models: A Bayesian Approach*

GARY KOOP †
Department of Economics, University of Edinburgh,
Edinburgh, U.K.
E-mail: G.Koop@ed.ac.uk

HERMAN VAN DIJK
Tinbergen Institute, Erasmus University,
Rotterdam, The Netherlands
E-mail: hkvdijk@tir.few.eur.nl

HENK HOEK
Tinbergen Institute, Erasmus University,
Rotterdam, The Netherlands
E-mail: hoek@tir.few.eur.nl

May 1997

ABSTRACT: In this paper, we make use of state space models to investigate the presence of stochastic trends in economic time series. A model is specified where such a trend can enter either in the autoregressive representation or in a separate state equation. Tests based on the former are analogous to Dickey-Fuller tests of unit roots, while the latter are analogous to KPSS tests of trend-stationarity. We use Bayesian methods to survey the properties of the likelihood function in such models and to calculate posterior odds ratios comparing models with and without stochastic trends. In addition, we extend these ideas to the problem of testing for integration at seasonal frequencies and show how are techniques can be used to carry out Bayesian variants of HEGY test or the Canova-Hansen test.

Keywords: State space models, Bayes Factor, Gibbs sampler,
unit root, seasonality

*Helpful comments from Carmen Fernandez, Philip Hans Franses, Richard Paap, Dale Poirier, Neil Shephard, Mark Steel and Yvo Steyn are gratefully acknowledged

†Financial support from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged

1 Introduction

State space models have been widely used in many fields in the physical and social sciences (see, eg., Harvey 1989). Such structural time series models can be used to analyze stochastic trends in macroeconomic and financial data.[1] In this paper, we use state space models and Bayesian methods to investigate whether stochastic trends are present in economic time series. In classical econometrics, a large number of tests have been developed which test for stochastic trends (see the survey by Stock 1994 or see Dickey and Fuller 1979). The vast majority of these tests have the unit root as the null hypothesis. In light of the low power of unit root tests, Kwiatkowski, Phillips, Schmidt and Shin (1994) developed a test for trend-stationarity, hereafter the KPSS test (ie. the null is trend stationarity and the alternative is the unit root).

The two types of classical tests can be illustrated in the following models. Dickey-Fuller type unit root tests use:

$$y_t = \rho y_{t-1} + e_t, \tag{1}$$

where e_t is a stationary error term and the null hypothesis is $\rho = 1$. A simple version of the KPSS test for stationarity makes use of a state space representation:

$$\begin{aligned} y_t &= \tau_t + e_t \\ \tau_t &= \tau_{t-1} + u_t, \end{aligned} \tag{2}$$

where u_t is white noise with variance σ_u^2, e_t is white noise with variance σ_e^2 and u_t and e_s are independent for all s and t. The null hypothesis is $\sigma_u^2 = 0$, in which case the series is stationary.

To our knowledge, Bayesian analysis of nonstationarity (see, among many others, DeJong and Whiteman 1991, Koop 1992, Phillips 1991 and Schotman and van Dijk 1991a,b) has focussed exclusively on generalizations of (1). Hence, one purpose of this paper is to develop Bayesian tests based on extensions of

[1] The reader is referred to Stock and Watson (1988) for an expository survey of stochastic trend behaviour in economic time series. One of the models they focus on is an unobserved components model which is a type of state space model.

(2) which can be used to test for stochastic trends by looking at σ_u^2 (as in the KPSS test) or by looking at the autoregressive coefficients (as in the Dicky-Fuller test) or both. The first part of this paper is devoted to analyzing evolving trends models (ie. investigating roots at the zero frequency). We begin by focussing on (2) to provide intuition into this class of models (see also Koop and van Dijk, 1996). For empirical relevance, however, it is important to allow for deterministic components and more general stationary dynamics. These are added as we generalize the model. The second part of the paper focusses on testing for integration at the seasonal frequency using the extension of (2) referred to as the evolving seasonals model (Hylleberg and Pagan 1995). In the context of seasonal models one can test for roots by looking at the autoregressive coefficients (see Hylleberg, Engle, Granger and Yoo 1990 — hereafter HEGY) or at coefficients similar to σ_u^2 (see Canova and Hansen 1995). We show how the evolving seasonals model can be used to nest both these approaches and, hence, Bayesian tests for seasonal integration analagous to HEGY or Canova-Hansen can be developed.

A further purpose of this paper is to develop computational tools for analyzing such models from a Bayesian perspective. For the models used in this paper, analytical results do not exist for posterior moments and posterior odds ratios. Hence, we must turn either to deterministic (like Gaussian product roles) or simulation-based integration techniques. The former are typically more accurate and faster in low dimensional problems while the latter are more suitable for high dimensional problems. For the models used in the first part of the paper, we show how (after analytically integrating out nuisance parameters) the problems can be reduced to one or two dimensions and, hence, numerical integration can be done using a deterministic technique. This obviates the need for sophisticated simulation methods or the direct use of Kalman filtering techniques. In the second part of the paper, it proves impossible to reduce the dimensionality of the problem below four or five. Hence, we turn to Gibbs sampling techniques to calculate posterior properties of the evolving seasonals models.

Throughout, we apply our techniques to simulated or macroeconomic data. We consider, in particular, the extended Nelson-Plosser data set and a well-known U.K. seasonal data set.

2 Testing for Integration in Evolving Trends Models

2.1 The Local Level Model

In this section, we begin with the simplest state space model given in (2) with the further assumptions that the errors, u_t and e_t are Normally distributed and that $\tau_0 = 0$. This model is referred to by Harvey (1989) as the local level model. Notice that there are several different ways of interpreting this model. First, it can be interpreted as saying that the observed series is decomposed into a local level plus error where the local level contains a unit root. Secondly, it can be interpreted as a time-varying parameter model (ie. τ_t is a mean which varies over time). Thirdly, by substituting the state equation into the measurement equation, the observed series can be seen to have an ARIMA(0,1,1) representation.

The local level model is parameterized in terms of σ_e^2 and σ_u^2. The latter of these parameters is crucial for testing. It proves convenient for both computation and prior elicitation to work in terms of a different parameterization: σ_e^2 and θ where:

$$\theta = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}.$$

The parameter θ has a simple intuitive interpretation. Furthermore, since θ is bounded between 0 and 1, a plausible noninformative prior is $p(\theta) = 1$ which is proper. Since Bayes factor calculations require proper priors on the parameter(s) being tested, the θ-parameterization allows us to calculate Bayes factors without subjectively eliciting informative priors. Of course, other priors could be used and the basic message of the current paper would not be altered. Further justification and discussion of priors is given in the Appendix. Formally, we assume the following prior:

$$p(\sigma_e^2, \theta) \propto \frac{1}{\sigma_e^2}$$

for $0 \leq \theta < 1$.

To develop a Bayesian version of the KPSS test, consider the Bayes factor (B_{01}) comparing $H_0 : \theta = 0$ to $H_1 : \theta > 0$, which can be calculated using the Savage-Dickey density ratio (see Verdinelli and Wasserman 1995). The Bayes factor can be written as:

$$B_{01} = \frac{p(\theta = 0|Data)}{p(\theta = 0)},$$

where the numerator of the Bayes factor is the marginal posterior of θ for the unrestricted model (or the alternative hypothesis) and the denominator is the marginal prior for θ evaluated at the point of interest $\theta = 0$ (or the null hypothesis).

For the local level model we can calculate the Savage-Dickey density ratio by integrating out the nuisance parameter σ_e^2.We set presample values of u_t to zero. By successively substituting the state equation into the measurement equation of the local level model we obtain:

$$y_t = e_t + \sum_{i=1}^{t} u_i,$$

and, hence, defining $y = (y_1, ..., y_T)'$,

$$y \sim N(0, \sigma_e^2 V),$$

where $V = I_T + \frac{\theta}{1-\theta} CC'$ and

$$C = \begin{pmatrix} 1 & 0 & . & . & . & . & 0 \\ 1 & 1 & 0 & . & . & . & 0 \\ 1 & 1 & 1 & 0 & . & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

C is known as the random walk generating matrix. Multiplying prior by likelihood and integrating out σ_e^2 yields the marginal posterior for θ :

$$p(\theta|Data) \propto |V|^{-\frac{1}{2}} (y'V^{-1}y)^{-\frac{T}{2}}. \tag{3}$$

The integrating constant of this posterior is, to our knowledge, not known in terms of elementary functions (such as the Gamma function). However, one-dimensional integration suffices to calculate it. If we combine this expression with the formula for the Savage-Dickey density ratio we obtain:

$$B_{01} = \frac{(y'y)^{-\frac{T}{2}}}{\int_0^1 |V|^{-\frac{1}{2}}(y'V^{-1}y)^{-\frac{T}{2}} d\theta}.$$

Hence, the Bayes factor with the uniform prior reduces to something similar to a likelihood ratio (with σ_e^2 integrated out), except the denominator of the likelihood ratio is an average over the parameter space under the alternative hypopthesis.

To illustrate our test procedure, we simulated two data sets from the local level model.[3] In all cases, $T = 100$ and $\sigma_e^2 = 1$. For the first data set $\theta = 0$ and for the second $\theta = 0.5$. Using simple numerical integration, we calculate the integrating constant for $p(\theta|Data)$ used in the Bayes factor. The marginal posteriors for θ for the two datasets are plotted in Figure 1. These marginal posteriors are quite reasonable. The Bayes factor comparing the stationary to the unit root model for the two data sets are 90.82 and 2.86×10^{-86}, respectively, indicating that they distinguish well between the two hypotheses.

A third data set is simulated from the standard AR(1) unit root model: $\Delta y_t = \varepsilon_t$, where ε_t is $i.i.N(0,1)$. Note that this model can be obtained from the local level by setting $\sigma_e^2 = 0$ and, hence, $\theta = 1$.[4] The resulting marginal posterior for θ is plotted in Figure 1. The Bayes factor in favour of stationarity is 9.85×10^{-146}. This suggests that if there is an AR unit root in the data generating process, our methods will be good at detecting it.

2.2 Adding an AR(1) Component

The Bayes factor above compares a white noise model to one with a random walk plus noise. With macroeconomic series, we are usually interested in testing whether a series can be characterized by stationary fluctuations around a deterministic trend, or whether it is better characterized by a stochastic trend. As a step in this direction, and as a way of illustrating the connections between the Dickey-Fuller and KPSS tests, consider:

[3] For all simulated data sets in the paper, we used the same seed for the random number generator.

[4] Note that when $\theta = 1$, the matrix V becomes infinite. Hence, formally speaking, the pure random walk model is not nested in the local level model, although the latter can come arbitrarily close to the former. This is why we restrict θ to lie in the interval $[0,1)$. When doing numerical integration we use a grid over the interval $[0,0.9999]$.

$$\begin{aligned} y_t &= \tau_t + \rho y_{t-1} + e_t \\ \tau_t &= \tau_{t-1} + u_t, \end{aligned} \tag{4}$$

where the assumptions about the errors are as in the previous section. If $\theta > 0$ and $|\rho| < 1$, then y_t has a random walk component plus a stationary component. If $\theta = 0$, then we get an AR(1) model.

We use the same prior as before with the additional assumption that $p(\rho)$ is uniform over the interval $[-1, 1]$ and ρ is *a priori* independent of the other parameters. If we condition on the initial observation, set presample values of u_t to zero, multiply likelihood function by prior and integrate out σ_e^2 analytically, we obtain:

$$p_1(\theta, \rho | Data) \propto |V|^{-\frac{1}{2}} [(y - \rho y_{-1})' V^{-1} (y - \rho y_{-1})]^{-\frac{T}{2}}, \tag{5}$$

where $y = (y_2, ..., y_T)'$ and $y_{-1} = (y_1, ..., y_{T-1})'$. Note that, if we had assumed an untruncated uniform prior for ρ, we could also have integrated out ρ analytically, using the properties of the Student-t density. Details are omitted here, see further below after equation (8). If we were to integrate out ρ, we could derive an expression for the Bayes factor analogous to that given above:

$$B_{01} = \frac{(y_{-1}' y_{-1})^{-\frac{1}{2}} (y' M y)^{-\frac{T-1}{2}}}{\int_0^1 |V|^{-\frac{1}{2}} (y_{-1}' V^{-1} y_{-1})^{-\frac{1}{2}} (s^2)^{-\frac{T-1}{2}} d\theta},$$

where $M = I - y_{-1}' (y_{-1}' y_{-1})^{-1} y_{-1}$ and $s^2 = (y - \widehat{\rho} y_{-1})' V^{-1} (y - \widehat{\rho} y_{-1})$. Furthermore, $\widehat{\rho} = (y_{-1}' V^{-1} y_{-1})^{-1} y_{-1}' V^{-1} y$.

In (4), a unit root is present if either $\theta > 0$ or $\rho = 1$. Formally, we consider four hypotheses:

$H_1 : \theta = 0$ and $|\rho| < 1$. In this case the series is stationary.

$H_2 : \theta > 0$ and $|\rho| < 1$. In this case the series is I(1) plus a stationary component.

$H_3 : \theta = 0$ and $|\rho| = 1$. In this case the series is I(1) and a random walk.

$H_4 : \theta > 0$ and $|\rho| = 1$. In this case the series is I(2).

Bayes factors for comparing these models can be calculated using (5) in the Savage-Dickey density ratio as described in Section 2.1. We label B_θ, B_ρ and $B_{\theta\rho}$ the Bayes factors for testing $\theta = 0$, $|\rho| = 1$ and

($\theta = 0, |\rho| = 1$), respectively. Although the setup here is more general than the simple Dickey-Fuller or Schotman and van Dijk (1991a,b) setup,[5] the similarities between B_ρ and these tests are apparent. The similarity between B_θ and the KPSS test is also apparent. However, our setup allows for for more general comparisons, involving joint tests. In fact, the posterior probability of any of the four hypotheses listed above can be calculated using B_θ, B_ρ and $B_{\theta\rho}$.

To investigate posterior properties and the performance of Bayesian model comparison procedures, we simulate data assuming $T = 100$ and $\sigma_e^2 = 1$. Table 1 presents posterior model probabilities for the four hypothesis listed above for different values of θ and ρ.

Table 1: Posterior Model Probabilities for Simulated Data Sets

	$p(H_1\|Data)$	$p(H_2\|Data)$	$p(H_3\|Data)$	$p(H_4\|Data)$
$\theta = 0, \rho = 0$	0.975	0.025	0.000	0.000
$\theta = .5, \rho = 0$	0.000	0.998	0.002	0.000
$\theta = 0, \rho = 1$	0.154	0.040	0.798	0.009
$\theta = .5, \rho = 1$	0.000	0.010	0.000	0.990
$\theta = .5, \rho = .5$	0.008	0.199	0.772	0.020
$\theta = 0, \rho = .5$	0.987	0.013	0.000	0.000

Note that our simulated data sets exhibit a wide variety of behavior: from white noise, through stationary but persistent, to I(1), to I(2) series. It can be seen that the Bayes factors, as reflected in the posterior model probabilities, do detect the appropriate degree of integration with high probability. In general, they also seem to detect whether nonstationarity is entering through an AR unit root or through a non-degenerate random walk state equation. The only exception is the case $\theta = .5, \rho = .5$ where more weight is put on the AR unit root than we would expect.

Figures 2-7 plot the joint and marginal posteriors for ρ and θ for the various simulated data sets. All results look completely reasonable. For instance, for the data set simulated with $\theta = \rho = 0$, the relevant posteriors (see Figure 2) are concentrated near these points. An examination of the joint posteriors indicates, unsurprisingly, that there tends to be a fair degree of correlation between θ and ρ. This correlation is most marked in the case of data simulated with $\theta = 0.5$ and $\rho = 0.5$ (see Figure 6) and accounts for the fact that the Bayes factors indicate that an AR unit root is present when in reality the unit root enters though the state equation. That is, the resulting marginals (Figures 6b and 6c) are very flat and,

[5] These authors would assume $e_t = 0$ for all t.

thus, allocate non-negligeable weight to the boundary points $\theta = 0$ and $\rho = 1$. The former increases the probability of $\theta = 0$ and the latter boosts $\rho = 1$. Combining these two we get more evidence for an AR unit root and less evidence for $\theta > 0$ than we would expect. Nevertheless, it should be emphasized that our approach does choose the correct degree of integration.

2.3 Extensions: More Dynamics and Deterministic Terms

Economic time series typically have more deterministic terms than (4) allows for. These considerations suggest that the following specification is more appropriate for empirical research:

$$\begin{aligned} \phi(L)y_t &= \tau_t + e_t \\ \tau_t &= \alpha + \tau_{t-1} + u_t \end{aligned} \tag{6}$$

where the assumptions about the errors are the same as for the previous models, but here we no longer assume $\tau_0 = 0$. It is worthwhile briefly motivating this particular extension as opposed to one which puts the deterministic component directly in the measurement equation or puts the AR component in the state equation. If we assume that $\phi(L)$ satisfies the stationarity conditions and difference y_t, we can write:

$$\phi(L)\Delta y_t = \alpha + u_t + \Delta e_t.$$

That is, if $\theta > 0$ the model becomes an ARIMA(p,1,1) plus drift. If $\theta = 0$, then the model can be written in terms of stationary fluctuations around a deterministic trend:

$$\phi(L)y_t = \tau_0 + \alpha t + e_t.$$

Hence, if we test $\theta = 0$ we are testing a null of trend-stationarity against an alternative of a unit root with drift. We feel that these are the sensible hypotheses to be considering in practice. An alternative way of extending (4) is less satisfactory. For instance, if we add the AR component to the state equation then, under $\theta = 0$, the model would reduce to white noise fluctations around a deterministic trend which is not a reasonable null hypothesis for most macroeconomic data.

Since we use numerical integration techniques in this section, it is important to reduce the dimensionality of the problem by integrating out nuisance parameters. It is convenient to rewrite the measurement equation in (6) as:

$$y_t = \tau_t + \rho y_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta y_{t-i} + e_t.$$

With this specification, we can focus on the bivariate posterior for θ and ρ in order to make inferences about the presence of stochastic trends.

By repeatedly substituting the state equation into the measurement equation in (6) we can write:

$$y_t = x_t \beta + v_t,$$

where $x_t = (y_{t-1}, 1, t, \Delta y_{t-1}, ..., \Delta y_{t-p+1}), \beta = (\rho, \gamma')', \gamma = (\tau_0, \alpha, \phi_1, ..., \phi_{p-1})', k = p+2$ and

$$v_t = e_t + \sum_{i=1}^{t} u_i.$$

Defining $y = (y_1, ..., y_T)\prime$, $X = (x_1', ..., x_T')'$ and treating p initial values of y_t as fixed[6] we obtain:

$$y \sim N(X\beta, \sigma_e^2 V).$$

Using the same prior as in the previous section plus untruncated uniform priors for the new parameters added in this section, and integrating out σ_e^2, we obtain an expression for the joint posterior of θ and β:

$$p_1(\theta, \beta | Data) \propto |V|^{-\frac{1}{2}} [(y - X\beta)' V^{-1} (y - X\beta)]^{-\frac{T}{2}}, \tag{7}$$

which is similar to (6).

To get the bivariate posterior for θ and ρ, we can integrate out γ using the presence of a Student-t kernel in (7), yielding:

$$p_1(\theta, \rho | Data) \propto |V|^{-\frac{1}{2}} |X^{*\prime} V^{-1} X^*|^{-\frac{1}{2}} s^{2^{-\frac{v}{2}}}, \tag{8}$$

[6]Note that, when we condition on p initial values, we are implicitly redefining T so that it is now equal to the old $T - p$. That is, we are treating our observed data as running from period $1 - p$ through T instead of as running from 1 through T as before. We maintain this convention throughout the remainder ot the paper.

where $\upsilon = T - k + 1$, X^* has $t'th$ row given by $x_t^* = (1, t, \Delta y_{t-1}, ..., \Delta y_{t-p+1})$,

$$s^2 = \frac{(y^* - X^*\widehat{\gamma})'V^{-1}(y^* - X^*\widehat{\gamma})}{\nu},$$

y^* has $t'th$ element given by $y_t^* = y_t - \rho y_{t-1}$ and $\widehat{\gamma} = (X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-1}y^*$. Using two-dimensional numerical integration we can calculate posterior properties of θ and ρ using equation (8). Bayes factors for the various hypothesis listed in the previous section can be calculated using the Savage-Dickey density ratio as described above.

Since this specification is now suitable for working with macroeconomic time series, in this section we investigate the properties of the extended Nelson-Plosser data in an empirical illustration. Schotman and van Dijk (1991b) use this data set to to carry out Bayesian tests for a unit root in an AR process (allowing for deterministic time trend). The reader is referred to this paper for a description of the data. In an attempt to make our results comparable to Schotman and van Dijk (1991b), we set $p = 3$ for all series except the unemployment rate for which we set $p = 4$. Table 2 presents posterior model probabilities for these series, the last column of this table presents the probability of a unit root calculated by Schotman and van Dijk.[7]

[7] The last column of Table 2 is taken from Hoek (1997), who made some corrections to Schotman and van Dijk's original calculations.

Table 2: Posterior Model Probabilities for Nelson-Plosser Data

	$p(H_1\|Data)$	$p(H_2\|Data)$	$p(H_3\|Data)$	$p(H_4\|Data)$	$S.v.D.$ $p(\rho=1)$
Real GNP	0.169	0.819	0.012	0.000	0.300
Nominal GNP	0.010	0.931	0.055	0.004	0.619
GNP per capita	0.247	0.740	0.013	0.000	0.290
Industrial Production	0.293	0.686	0.021	0.000	0.316
Employment	0.002	0.998	0.001	0.000	0.313
Unemployment	0.463	0.533	0.004	0.000	0.217
GNP Deflator	0.011	0.866	0.110	0.014	0.678
Consumer Prices	0.000	0.996	0.003	0.001	0.697
Nominal Wages	0.026	0.887	0.078	0.010	0.602
Real Wages	0.006	0.948	0.042	0.004	0.642
Money	0.036	0.897	0.055	0.012	0.397
Velocity	0.001	0.983	0.015	0.000	0.666
Interest Rate	0.001	0.973	0.011	0.015	0.641
Stock Prices	0.021	0.898	0.079	0.001	0.653

The results in Table 2 accord reasonably well with the results of Schotman and van Dijk (1991b), despite differences in specification (and slight differences in the prior). In particular, most evidence for stationarity is found for series like real GNP, GNP per capita, unemployment and industrial production. Other series provide much stronger evidence of integration. The present approach, however, finds more evidence of nonstationarity. One possible explanation for this is that it could be due the difference in treatment of deterministic terms. Schotman and van Dijk treat these in a manner that is different from the present approach. Another possibility is that the implicit MA component added in our state-space approach is an important extension for macro data. It is interesting to note that, for most series, H_2 receives much more probability than H_3 indicating that the data prefer the state space unit root (which implicitly adds a moving average component) to the autoregressive unit root. To see why this might increase the probability of integration, suppose that a true data generating process exists and its is ARIMA(3,1,1) and that the MA coefficient is substantial and negative. This series, of course, is I(1) and we would hope a test would indicate this. The Schotman and van Dijk approach would approximate the ARIMA(3,1,1) by an AR(3) model. The presence of a negative MA coefficient would tend to pull the AR coefficients into the stationary region, reducing the probability of the unit root relative to the present approach which would correctly model the ARIMA(3,1,1).

3 Testing for Integration in the Evolving Seasonals Model

3.1 Theory

The evolving seasonals model has recently been reintroduced to the econometrics literature in Hylleberg and Pagan (1995). Originally developed in Hannan, Terrell and Tuckman (1970), this model is a very flexible specification which allows the seasonal pattern to vary over time. A simple variant of this model is given by:

$$y_t = \tau_{0t}\cos(\alpha_0 t) + \tau_{1t}\cos(\alpha_1 t) + 2\tau_{2t}\cos(\alpha_2 t) + 2\tau_{3t}\sin(\alpha_2 t) + e_t, \tag{9}$$

where $\alpha_0 = 0, \alpha_1 = \pi$ and $\alpha_2 = \frac{\pi}{2}$ capture behaviour at the relevant 0 and seasonal frequencies, respectively. The $\tau_{it}'s$ capture the evolution of the trend and seasonal patterns over time. Hylleberg and Pagan (1995) shows how this specification nests most common seasonal models.

In this paper we focus on testing for seasonal unit roots so it is worthwhile to briefly digress and describe the two chief classical approaches. The most common of these is outlined in Hyllegerg, Engle, Granger and Yoo (1990) — HEGY — and is based on the fact that an AR(p) specification: $\phi(L)y_t = e_t$ can be written as

$$\phi^*(L)y_{4,t} = \delta_0 y_{1,t-1} + \delta_1 y_{2,t-1} + \delta_2 y_{3,t-2} + \delta_3 y_{3,t-1} + e_t,$$

where $y_{1,t} = (1 + L + L^2 + L^3)y_t, y_{2,t} = -(1-L)(1+L)y_t, y_{3,t} = -(1-L^2)y_t$ and $y_{4,t} = (1-L^4)y_t$. A nonseasonal unit root is present if $\delta_0 = 0$, while if $\delta_1 = 0$ a seasonal unit root at frequency π is present. δ_2

and δ_3 relate to possible seasonal unit roots at frequency $\frac{\pi}{2}$ and HEGY suggests a joint test of $\delta_2 = \delta_3 = 0$. An alternative test is given by Canova and Hansen (1994) and is based on a specification similar to (9) under the assumption that, for $i = 0, 1, 2, 3$:

$$\tau_{it} = \tau_{i,t-1} + u_{it},$$

and $var(u_{it}) = \sigma_i^2$. If $\sigma_1^2 = 0$ then a seasonal unit root at frequency π is present while if $\sigma_2^2 = \sigma_3^2 = 0$ then a seasonal unit root at frequency $\frac{\pi}{2}$ is present. The nonseasonal unit root occurs if $\sigma_0^2 = 0$.

Given the evolving seasonals model, it is apparent that we can derive a specification that nests both these approaches in the same way that our specification in the previous section nested both Dickey-Fuller and KPSS tests. As before, it is important to allow for deterministic terms and hence we work with the following specification:

$$\begin{aligned}
\phi^*(L)y_{4,t} &= \tau_{0t} + \tau_{1t}\cos(\pi t) + 2\tau_{2t}\cos(\frac{\pi t}{2}) + 2\tau_{3t}\sin(\frac{\pi t}{2}) \\
&\quad + \delta_0 y_{1,t-1} + \delta_1 y_{2,t-1} + \delta_2 y_{3,t-2} + \delta_3 y_{3,t-1} + e_t \\
\tau_{it} &= \alpha_i + \tau_{i,t-1} + u_{it},
\end{aligned} \tag{10}$$

where the $e_t' s$ are $i.i.N(0, \sigma_e^2)$, the $u_{it}' s$ are $i.i.N(0, \sigma_i^2)$ and all error terms are independent of one another. As in the previous section, we can test for unit roots either through the AR coefficients or through the error variances in the state equations (eg. testing $\delta_0 = 0$ or $\sigma_0 = 0$ for the nonseasonal unit root). If the state equations are substituted into the measurement equation it can be seen that the $\tau_{i0}' s$ enter as a deterministic seasonal pattern and the inclusion of drift terms in the state equations (ie. the $\alpha_i' s$) allows for a deterministic trend the specification and in the seasonal patterns. In our empirical work, we rule out the latter and set $\alpha_1 = \alpha_2 = \alpha_3 = 0$, but leave α_0 unrestricted. Assuming the AR coefficients satisfy the stationarity condition, then if $\sigma_i = 0$ for $i = 0, 1, 2, 3$ the model is characterized by stationary fluctations around a deterministic seasonal pattern. Hence, (10) is an extremely flexible specification which nests most common seasonal models, and our Bayesian counterpart to the Canova-Hansen test has as its null hypothesis a reasonable model for macroeconomic time series.

As before, we reparameterize in terms of

$$\theta_i = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_e^2}.$$

Tests of the various sorts of seasonal integration then reduce to testing for zero restrictions on the $\theta_i' s$.

Note, however, that there are eight parameters of interest (ie. δ_i and θ_i for $i = 0, 1, 2, 3$), so that, even if we analytically integrate out all nuisance parameters, deterministic numerical integration is extremely difficult given current computational power. However, it is possible to set up a Gibbs sampler to analyze this model. To calculate Bayes factors, it is necessary to specify priors for the $\theta_i' s$. To do this, we extend the strategy of the previous section, assume prior independence between these parameters, and obtain: $p(\theta_i) = 1$ if $0 \leq \theta_i < 1$.[8] For all other parameters, we use traditional, flat, noninformative priors. Hence, the Bayes factors calculated here have the same "weighted likelihood ratio" form as in the previous section. Of course, subjective informative priors can be used if so desired.

Conditional on knowing σ_e^2 and θ_i for $i = 0, 1, 2, 3$, the Gibbs sampler can be set up exactly as in de Jong and Shephard (1995).[9] In particular, our equation (10) is exactly in the form as the model in Section 3 of de Jong and Shephard if we condition on p initial observations. Using their equations (2) and (4) modified for the inclusion of regression effects as in their Section 5, we can sample jointly from all the states and all regression parameters jointly (conditional on σ_e^2 and the $\theta_i' s$). In our experience, the de Jong-Shephard algorithm is highly efficient. Of particular value is the fact that it reduces the Gibbs sampler to three blocks. For the sake of brevity, we do not repeat the exact form of the algorithm here, but refer the reader to de Jong and Shephard (1995).

The conditional distribution of σ_e^{-2} is:

$$p(\sigma_e^{-2} | Data, \phi^*, \delta_0, \delta_1, \delta_2, \delta_3, \tau) = f_G(\sigma_e^{-2} | \frac{T}{2}, \frac{\sum_{i=1}^T e_t^2}{2}), \tag{11}$$

[8]Note that we are using an improper prior for the $\delta_i' s$ and, hence, do not calculate Bayes factors for these parameters. The methodology outlined in this section could be used do this, but proper priors would be needed. Such priors could either be elicited subjectively or we could use a flat prior over the stationary region. The necessary restriction for imposing the latter is complicated (see Franses, 1996, pp. 64-66). Hence, for reasons of simplicity and to keep the empirical illustration focussed on the $\theta_i' s$, we do not consider proper priors for the AR parameters.

[9]Fruhwirth-Schnatter (1994) and Carter and Cohn (1994) provide alternative methods for Gibbs sampling with state space models.

where $f_G(.|a,b)$ is the Gamma density with mean $\frac{a}{b}$ and variance $\frac{a}{b^2}$. The posterior conditionals for the $\theta_i's$ (for $i = 0, 1, 2, 3$), can be obtained by noting that they are independent of one another (i.e. θ_i is only in the $i'th$ state equation) and are closely related to the variance of each equation. The resulting conditional posterior is non-standard:

$$p(\theta_i|Data, \sigma_e^2, \tau) \propto \left(\frac{1-\theta_i}{\theta_i}\right)^{\frac{T}{2}} \exp\left(-\frac{1-\theta_i}{\theta_i} SSE_i\right), \qquad (12)$$

where

$$SSE_i = \sum_{j=1}^{T} \frac{u_{ij}^2}{2\sigma_e^2}.$$

Hence, we use an alternative approach. If we had parameterized with $\lambda_i = \frac{\sigma_i^2}{\sigma_e^2}$ and used a flat prior for λ_i, then resulting posterior conditional for λ_i would be Gamma:

$$p(\lambda_i|Data, \sigma_e^2, \tau) \propto \lambda_i^{-\frac{T}{2}} \exp\left(-\frac{SSE_i}{\lambda_i}\right).$$

But a flat prior for λ_i is improper and has some strange implications for θ_i (see Appendix). The uniform prior for θ_i, which is truncated to ensure $0 \leq \theta_i < 1$, is proper and implies a prior for λ_i which is proportional to $\frac{1}{(1+\lambda_i)^2}$. This suggests as simple strategy for drawing from θ_i using a Metropolis-Hastings algorithm (see, for instance, Chib and Greenberg, 1995). Suppose the current draw of λ_i is called λ_i^{Old}. First take a candidate draw of λ_i from (12) using the Gamma distribution (call it λ_i^{New}). This draw is accepted with probability:

$$\frac{\frac{1}{(1+\lambda_i^{New})^2}}{\frac{1}{(1+\lambda_i^{Old})^2}},$$

where probabilities greater than one are rounded down to one. If the candidate draw is not accepted then the draw for λ_i remains λ_i^{Old}. Draws from λ_i can be converted into draws from θ_i using the fact that $\theta_i = \frac{\lambda_i}{1+\lambda_i}$.

Output from this posterior simulator can be used to calculate posterior features of interest as well as the Bayes factor using the Savage-Dickey density ratio (see, for instance, Verdenelli and Wasserman, 1995, section 2.2).[10]

3.2 Empirical Illustration

The techniques described above are here illustrated using several UK seasonal series: GDP, total consumption (TOTCON), consumption of nondurables (NONDUR), total investment (TOTINV), exports (EXPORTS) and imports (IMPORTS). All data are quarterly, logged and run from 1955:1 to 1988:4. These series have been analysed extensively by many authors (see Franses, 1996, chapter 5 for a list of citations). Franses, 1996, Table 5.2 presents results from the HEGY test on these series (and others), concluding that the nonseasonal unit root seems to be present in all series, and TOTCON and NONDUR have in addition roots at both seasonal frequencies.Table 3 presents Bayes factors for testing $\theta_i = 0$, which we call B_{θ_i} for $i = 0, 1, 2, 3$. Small values of B_{θ_i} indicate evidence in favour of seasonal integration. The last four rows present posterior means of the $\delta_i's$, with posterior standard deviations in parentheses.

Table 3: Posterior Information on UK Seasonal Series

	GDP	$TOTCON$	$NONDUR$	$EXPORTS$	$IMPORTS$	$TOTINV$
B_{θ_o}	$4.9x10^{-114}$	$1.5x10^{-13}$	$3.2x10^{-41}$	$7.0x10^{-142}$	$8.4x10^{-3}$	$2.1x10^{-35}$
B_{θ_1}	0.10	$5.8x10^{-3}$	$6.3x10^{-4}$	0.26	$5.7x10^{-2}$	0.27
B_{θ_2}	0.14	$2.4x10^{-2}$	$2.7x10^{-3}$	0.66	$2.8x10^{-2}$	$3.1x10^{-2}$
B_{θ_3}	0.31	$4.8x10^{-3}$	$7.4x10^{-4}$	0.18	0.15	0.29
δ_0	−0.19 (0.07)	−0.11 (0.05)	−0.09 (0.07)	−0.35 (0.07)	−0.20 (0.06)	−0.16 (0.05)
δ_1	−0.51 (0.21)	−0.75 (0.26)	−0.82 (0.34)	−0.27 (0.06)	−0.41 (0.10)	−0.34 (0.18)
δ_2	−0.53 (0.11)	−0.98 (0.23)	−0.77 (0.19)	−0.63 (0.09)	−0.69 (0.10)	−0.62 (0.12)
δ_3	−0.21 (0.12)	−0.34 (0.27)	−0.40 (0.25)	0.08 (0.08)	−0.21 (0.13)	−0.02 (0.14)

A standard Bayesian rule of thumb (see, eg., Poirier, 1995, page 380) is to say that there is slight evidence against $\theta_i = 0$ if $B_{\theta_i} > 0.10$, strong evidence if $0.01 \leq B_{\theta_i} \leq 0.10$, and decisive evidence if $B_{\theta_i} < 0.01$. Using this rule of thumb, all series provide decisive evidence in favour of a unit root at the nonseasonal frequency. TOTCON and NONDUR provide decisive evidence in favour of roots at both seasonal frequencies. These results accord with those provided by the HEGY test. The Bayes factors for

[10]Due to the difficulties of evaluating (12) at the point 0, we evaluate it at a point close to zero. Formally speaking, this means we are testing the hypothesis that $\theta_i = 0.0001$ rather than $\theta_i = 0$. In practical applications the differences between these two hypotheses are negligeable.

the seasonal unit roots for the other series do not provide decisive evidence, but nevertheless some evidence for seasonal unit roots occurs.

Our specification allows for seasonal and nonseasonal unit roots to enter through either the AR coefficients or the state equation. Although we do not calculate Bayes factors for the former, the posterior moments for the $\delta_i's$ indicate that the data chooses to put unit roots (if they exist) in the state equations. This finding is analagous to that noted in Section 2, where the Nelson-Plosser data tended to favor H_2 over H_3.

It is also worth noting that we test each of the $\theta_i's$ individually. Given the aliasing problem, one may be interested in doing a joint test of $\theta_2 = \theta_3 = 0$. This can, of course, be easily done using our present framework.

4 Conclusion

In this paper, we develop Bayesian tests of stochastic trends using state space representations. We consider both non-seasonal and seasonal models, and AR unit roots and roots arising in the state equation(s). Our general framework nests most of the common approaches to testing for integration in the literature. We construct computational methods involving either numerical integration or posterior simulation to calculate the probability associated with each type of integration. Empirical evidence using simulated and real data indicate that the approach advocated in this paper is both simple to use and yields reasonable results. The added flexibility of state space modelling and the allowance for the test of stationarity to be a point hypothesis (in contrast to the usual setup where the unit root is the point hypothesis) heighten the advantages of our approach.

The basic ideas in this paper can be extended in two important ways: i) Different priors can easily be accomodated; and ii) Multivariate models and resulting issues involving common trends can be handled in a conceptually similar manner.

5 Appendix: Priors and Parameterizations

In the local level model, we parameterize the variance of the state equation in terms of the parameter:

$$\theta = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

which has a natural interpretation. Here, and in the material below, we use the standard noninformative prior for σ_e^2, independent of the other parameter in the model. We use a flat prior over the interval $[0,1)$ and the stationary case corresponds to $\theta = 0$. Since the prior is finite at the latter point and proper, the Savage-Dickey density ratio can be used to calculate Bayes factors. This simplifies our computations and allows for the analytical derivation of the Bayes factor formulae in the first part of the paper. Of course, the general ideas in this paper hold for any prior.

The prior we use can be interpreted as a "noninformative" one. However, there are other candidates for a "noninformative" prior. An alternative is the standard noninformative prior for the variance:

$$p(\sigma_u^2) \propto \frac{1}{\sigma_u^2}.$$

This prior is improper and is infinite at the point $\sigma_u^2 = 0$. Hence, meaningful Bayes factors cannot be obtained and, even if the prior were truncated to be proper, the Savage-Dickey density ratio could not be used to calculate the Bayes factor. By using the change of variable theorem[11] it can be seen that this prior implies a prior for θ of the form:

$$p(\theta) \propto \frac{1}{(1-\theta)\theta},$$

which is a U-shaped prior with infinite asymptotes at the points $\theta = 0$ and $\theta = 1$.

Alternatively, it is common to parameterize the model in terms of $\lambda = \frac{\sigma_u^2}{\sigma_e^2}$. There are two common ways of being noninformative about λ. The first is used in Min (1992) and is given by $p(\lambda) \propto 1$. In previous work with this improper prior, we found significant evidence that it resulted in an improper posterior which is a

[11]Formally speaking, the change of variable theorem should not be used here since the prior is not a valid probability density function. Nevertheless, if we treat the resulting formulae as coming from a prior truncated at a very large value they can at least provide some intuitive insight into the implications of the various noninformative priors considered in this Appendix.

reason for avoiding it. Nevertheless, if one were to truncate it to make the prior (and, hence, the posterior) proper it can still be used for testing using the Savage-Dickey density ratio. Results using this prior are slightly sensitive to the choice of truncation point, but nevertheless reasonable choices of truncation point give results similar to those given in the body of the paper. In terms of our preferred parameterization, this flat prior for λ implies:

$$p(\theta) \propto \frac{1}{(1-\theta)^2}$$

indicating a prior which puts very large weight on values of θ near one. An alternative noninformative prior for λ might be $p(\lambda) \propto \frac{1}{\lambda}$. This prior suffers from the same problems for Bayes factor calculation as the noninformative prior for σ_u^2. In terms of our preferred parameterization, this prior implies:

$$p(\theta) \propto \frac{1-\theta}{\theta}$$

which puts very large weight on small values of θ.

Hence, we have four different "noninformative" priors which imply very different prior views about θ (ie. uniform, U-shaped, skewed towards 0 and skewed towards 1 on θ). This illustrates the great care that must be taken in prior elicitation, even when the researcher is striving to be noninformative. However, we have found that, for reasonably large sample sizes (e.g. $T > 50$) that the choice of noninformative prior has little effect on posterior inference. In a more serious empirical exercise, the researcher would likely have prior information which could be used to guide construction of a suitable informative prior.

Of course, the testing strategy discussed in the paper could be performed using any prior. In some cases, however, the Savage-Dickey density ratio could not be used and the researcher would be forced to use more computationally intensive methods (see, eg., Chib 1995) to calculate Bayes factors.

References

[1] Canova, F. and Hansen, B. (1994). "Are seasonal patterns constant over time? A test for seasonal stability," *International Economic Review.*

[2] Carter, C. and Kohn, R. (1994). "On Gibbs sampling for state space models," *Biometrika*, 81, 541-553.

[3] Chib, S. (1995). "Marginal likelihood from the Gibbs output," *Journal of the American Statistical Association*, 90, 1313-1321.

[4] Chib, S. and Greenberg, E. (1995). "Understanding the Metropolis-Hastings algorithm," *The American Statistician*, 49, 327-335.

[5] DeJong, D. and Whiteman, C. (1991). "The temporal stability of dividends and stock prices: Evidence from the likelihood function," *American Economic Review*, 81, 600-617.

[6] Dickey, D. and Fuller, W. (1979). "Distribution of the estimators for autoregressive series with a unit root," *Journal of the American Statistical Association*, 74, 427-431.

[7] Franses, P.H. (1996). *Periodicity and Stochastic Trends in Economic Time Series*, Oxford: Oxford University Press.

[8] Fruhwirth-Schnatter, S. (1994). "Data augmentation and dynamic linear models," *Journal of Time Series Analysis*, 15, 183-202.

[9] Hannan, E.J., Terrell, R. and Tuckwell, N. (1970). "The seasonal adjustment of economic time series," *International Economic Review*, 11, 24-52.

[10] Harvey, A. (1989). *Forecasting, Structural Time Series Models, and the Kalman Filter.* Cambridge: Cambridge University Press.

[11] Hoek, H. (1997). "Variable tends: A Bayesian perspective," *PhD dissertation, Tinbergen Institute, Erasmus University.*

[12] Hylleberg, S., Engle, R., Granger, C., and Yoo, B. (1990). "Seasonal integration and cointegration," *Journal of Econometrics*, 44, 215-238.

[13] Hylleberg, S. and Pagan, A. (1995). "Seasonal integration and the evolving seasonals model," *Australian National University, Working Papers in Economics and Econometrics,* no. 281.

[14] de Jong, P. and Shephard, S. (1995). "The simulation smoother for time series models," *Biometrika*, 82, 339-350.

[15] Koop, G. (1992). "'Objective' Bayesian unit root tests," *Journal of Applied Econometrics*, 7, 65-82.

[16] Koop, G. and van Dijk, H. (1996). "Testing for integration using evolving trends models," *Proceedings of the American Statistical Association, Section on Bayesian Statistics*, forthcoming.

[17] Kwiatkowski, D., Phillips, P.C.B, Schmidt, P. and Shin, Y. (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root," *Journal of Econometrics*, 54, 159-178.

[18] Min, C. (1992). "Unconditional estimation of time-varying parameter models: A Gibbs sampling approach," manuscript.

[19] Phillips, P.C.B. (1991). "To criticize the critics: An objective Bayesian analysis of stochastic trends," *Journal of Applied Econometrics*, 6, 333-364.

[20] Poirier, D. (1995). *Intermediate Statistics and Econometrics*, Cambridge: The MIT Press.

[21] Schotman, P. and van Dijk, H. (1991a). "A Bayesian analysis of the unit root in real exchange rates," *Journal of Econometrics*, 49, 195-238.

[22] Schotman, P. and van Dijk, H. (1991b). "On Bayesian routes to unit roots," *Journal of Applied Econometrics*, 6, 387-402.

[23] Stock, J. (1994). "Unit roots and trend breaks in econometrics," in *Handbook of Econometrics*, vol. 4, ed. by R. Engle and D. McFadden. New York: North-Holland, 274-2781.

[24] Stock, J. and Watson, M. (1988). "Variable trends in economic time series," *Journal of Economic Perspectives*, 2, 147-174.

[25] Verdinelli, I. and Wasserman, L. (1995). "Computing Bayes factors using a generalization of the Savage-Dickey density ratio," *Journal of the American Statistical Association*, 90, 614-618.

Figure 1: Marginal for Theta

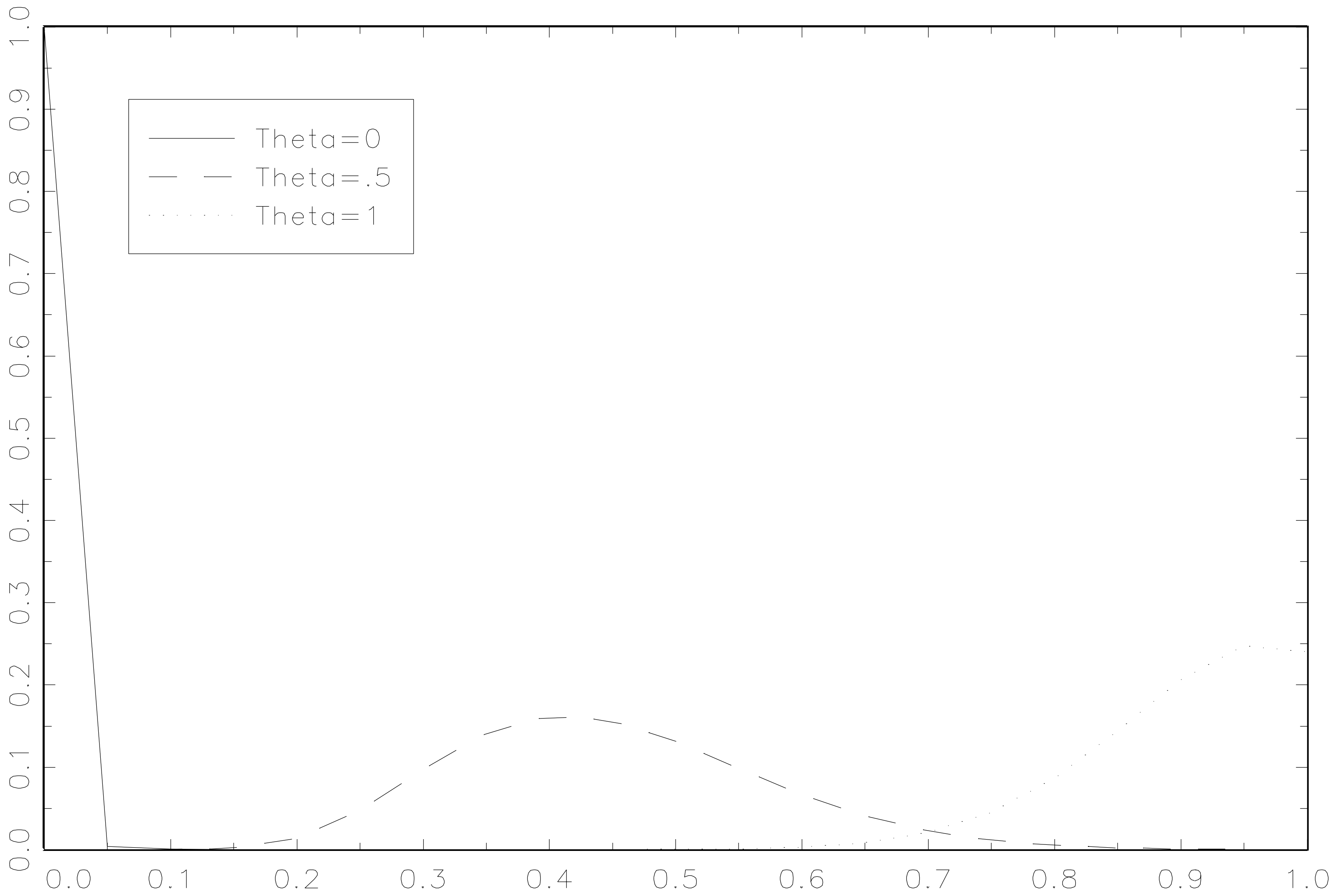

Figure 2a: Joint for Theta and Rho

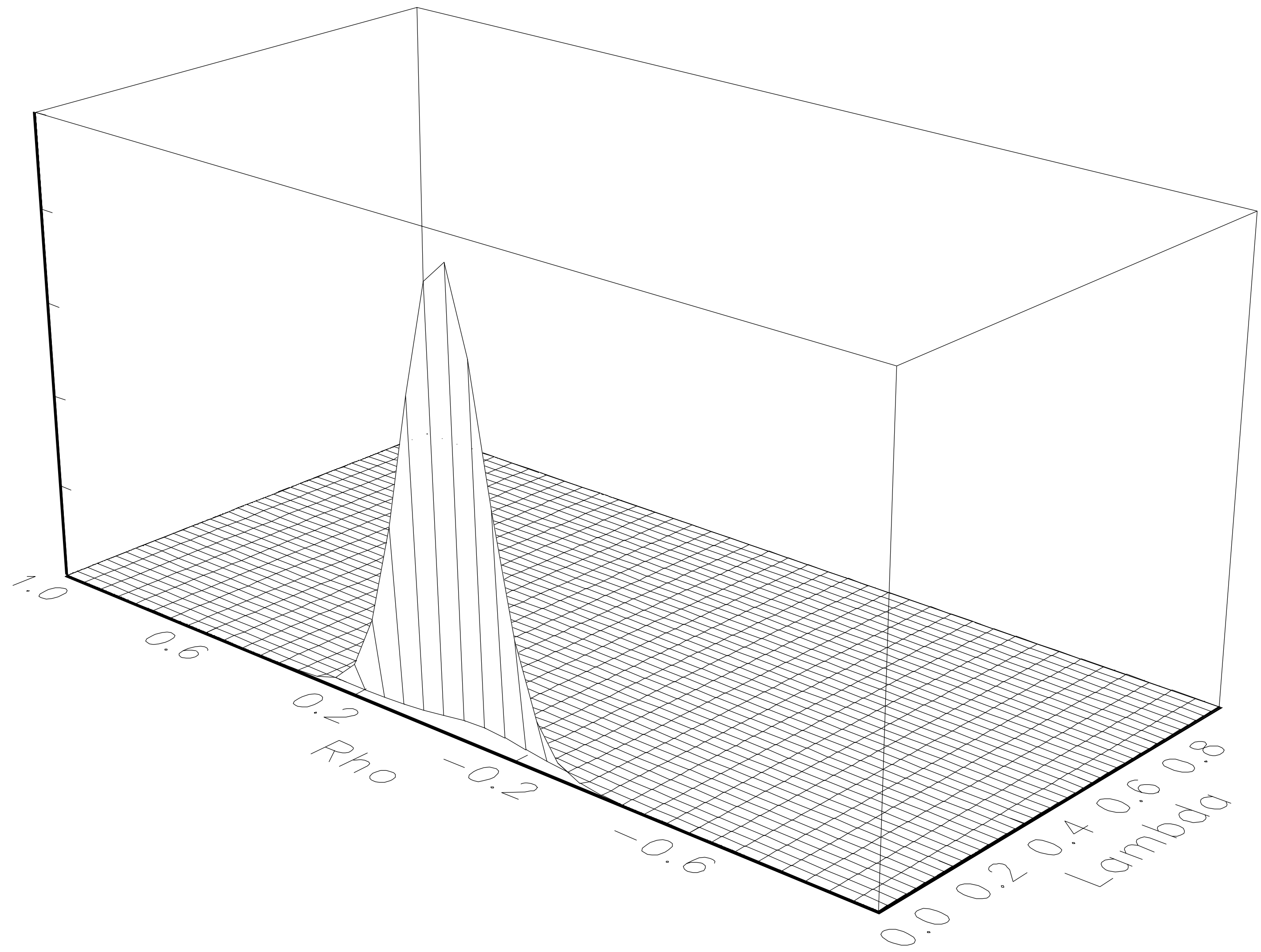

Figure 2b: Marginal for Theta

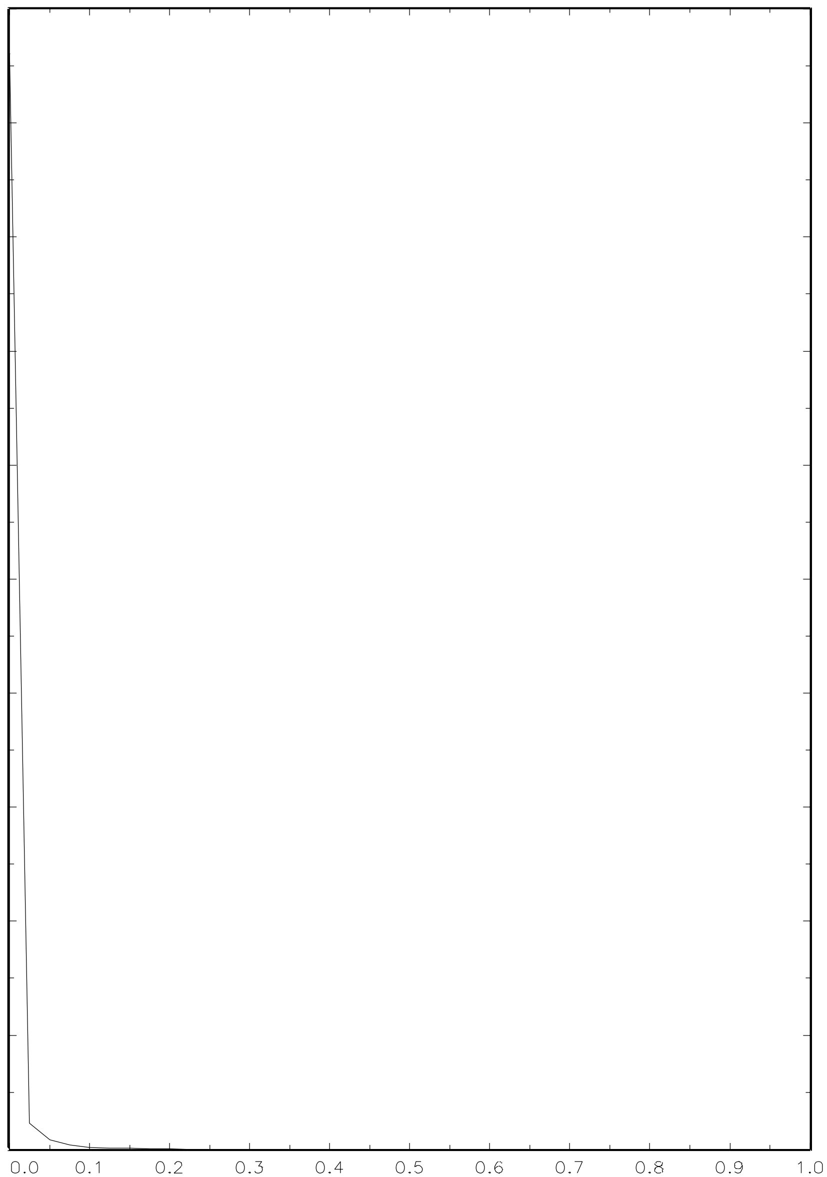

Figure 2c: Marginal for Rho

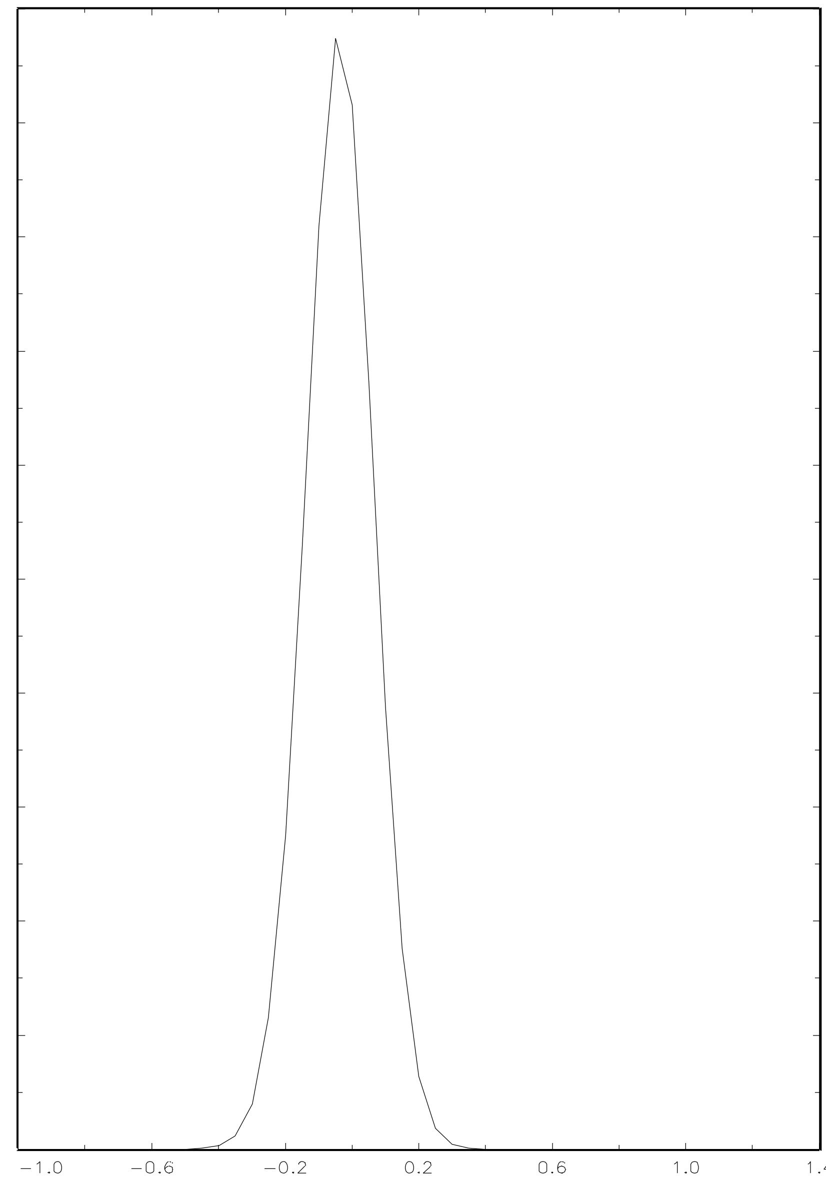

Figure 3a: Joint for Theta and Rho

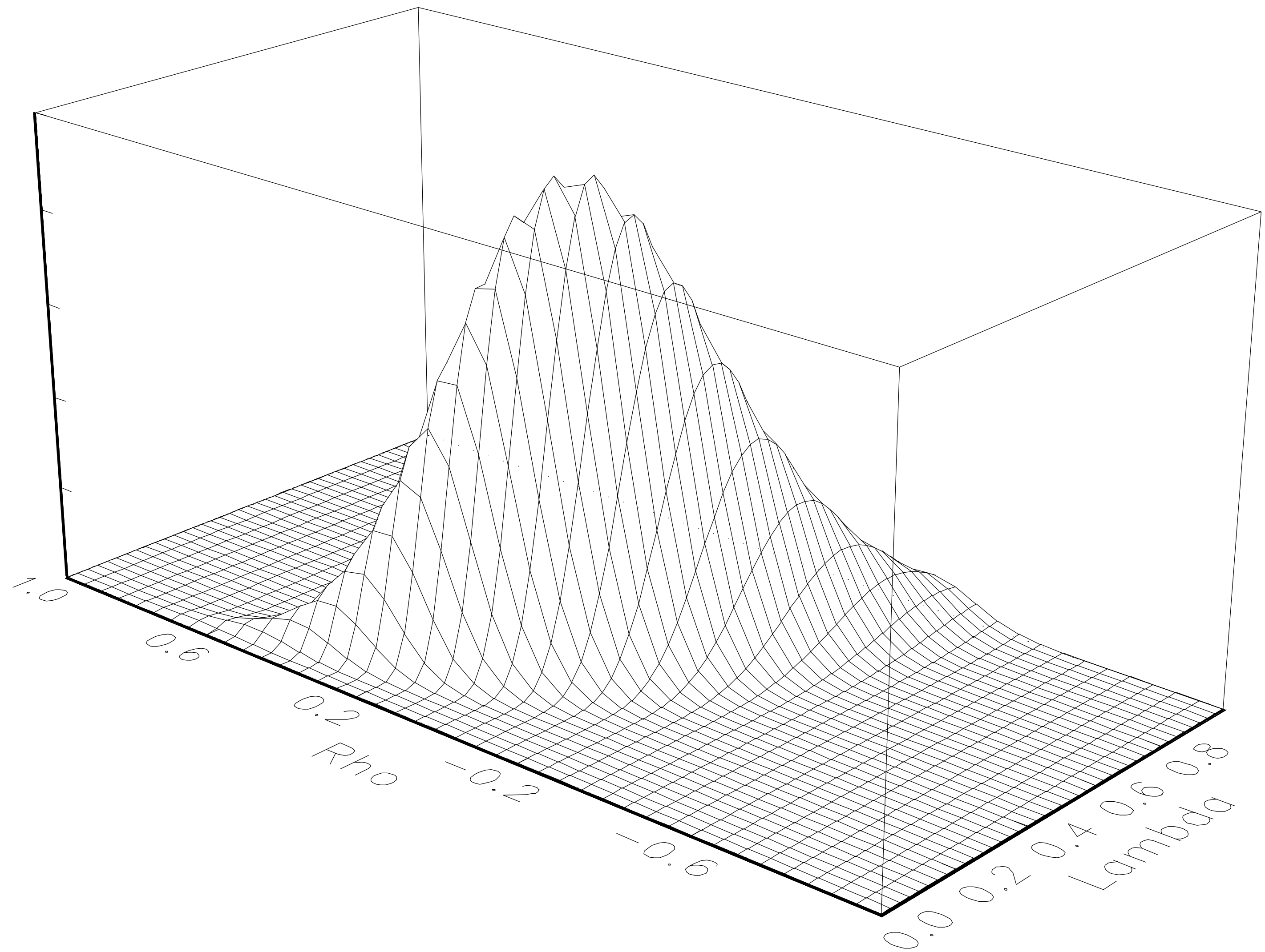

Figure 3b: Marginal for Theta

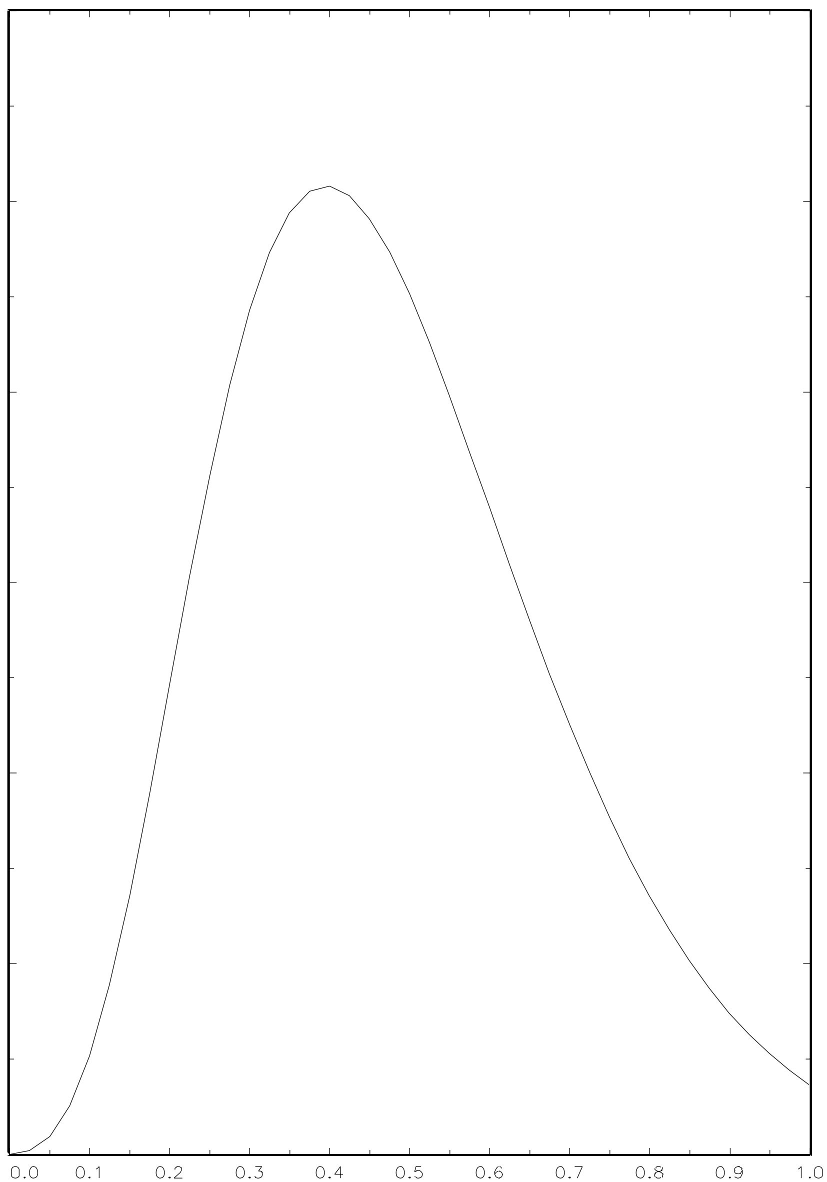

Figure 3c: Marginal for Rho

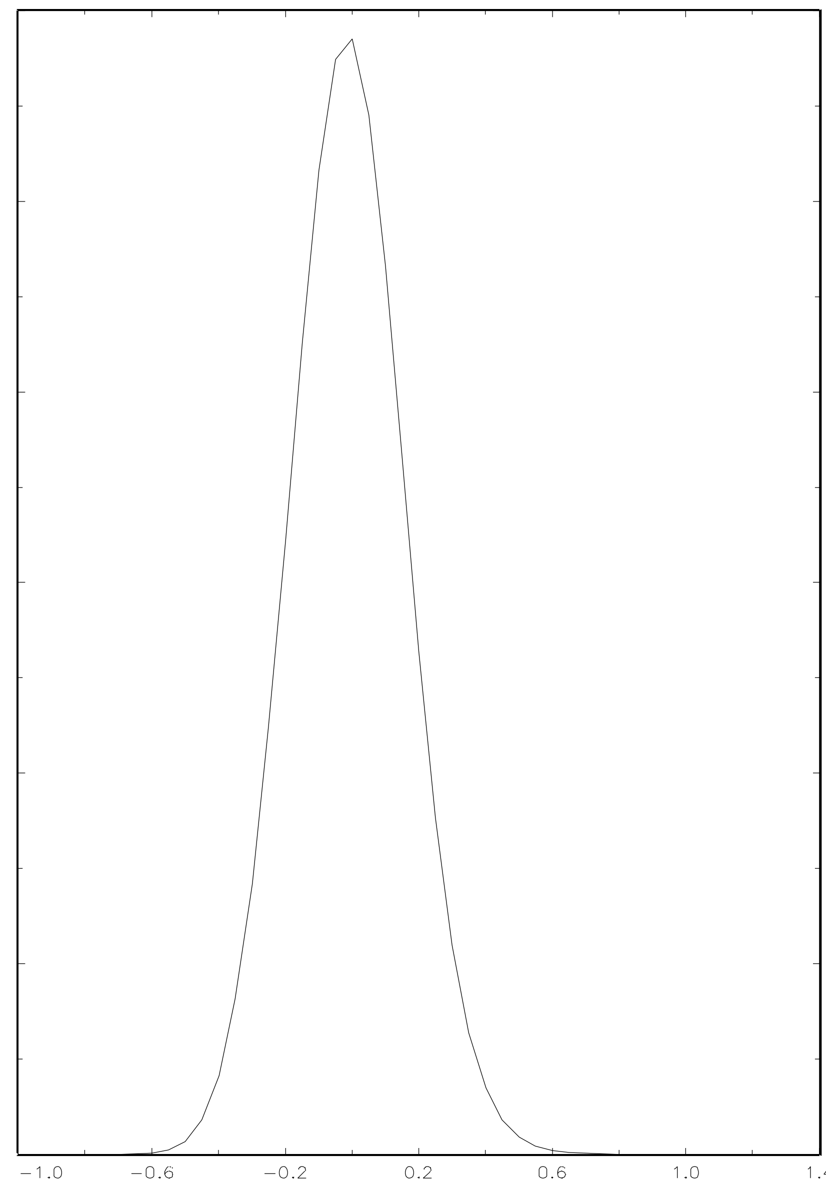

Figure 4a: Joint for Theta and Rho

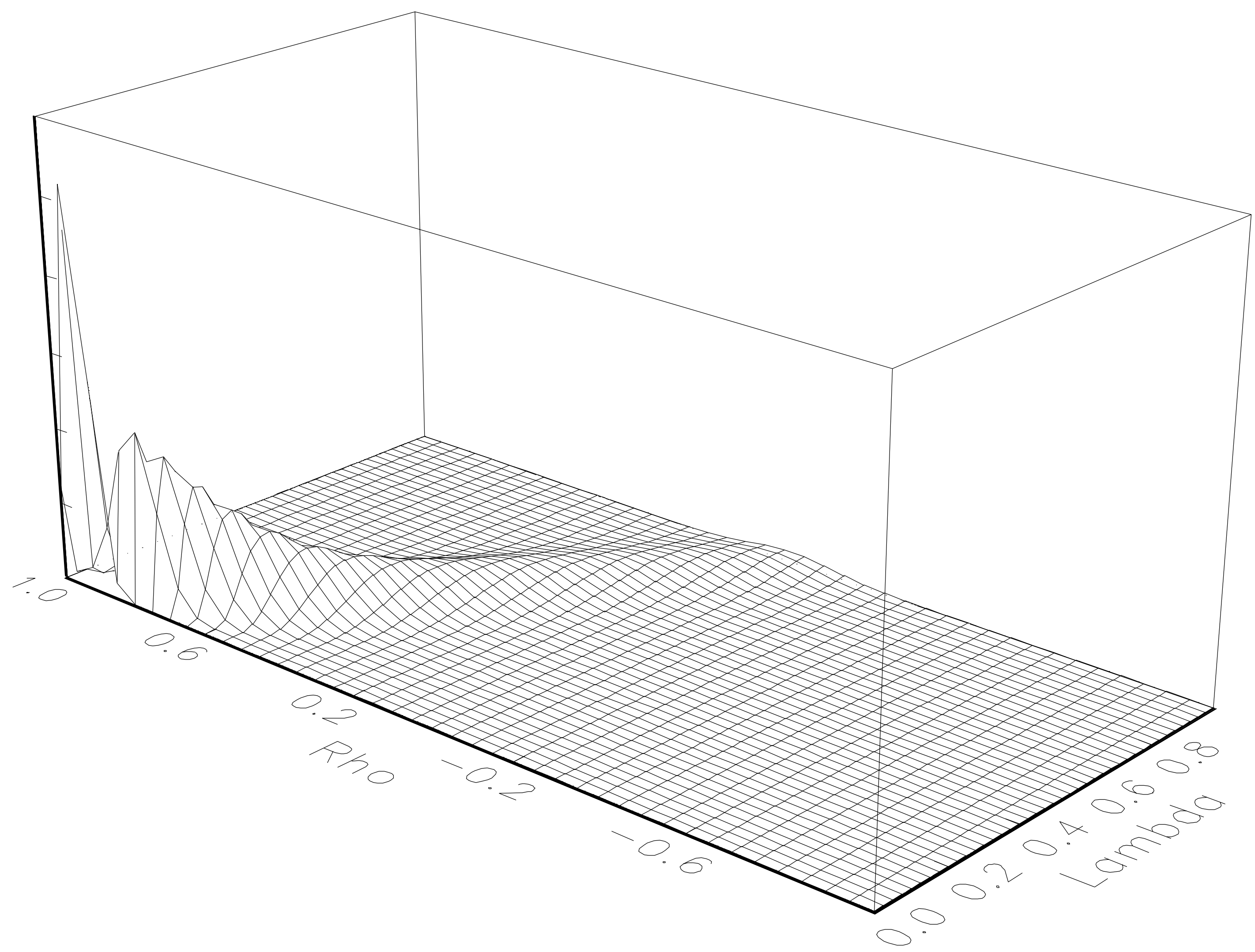

Figure 4b: Marginal for Theta

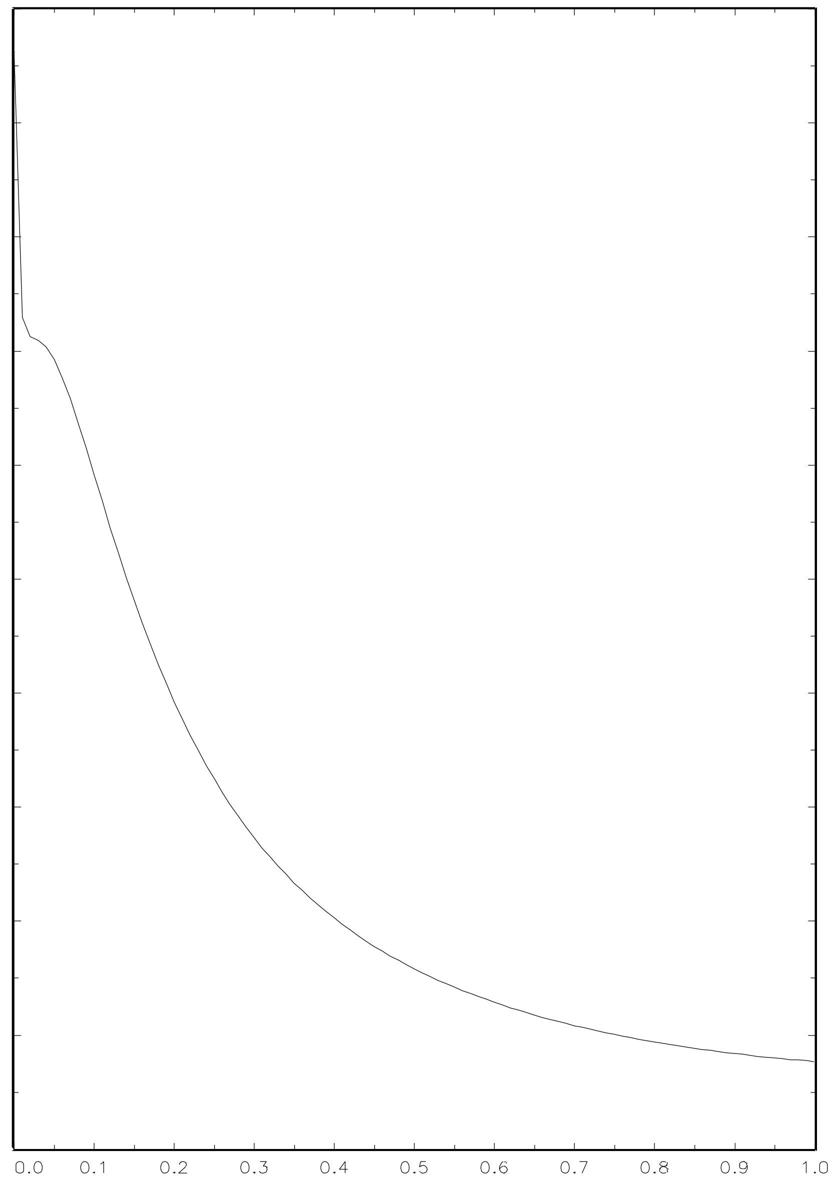

Figure 4c: Marginal for Rho

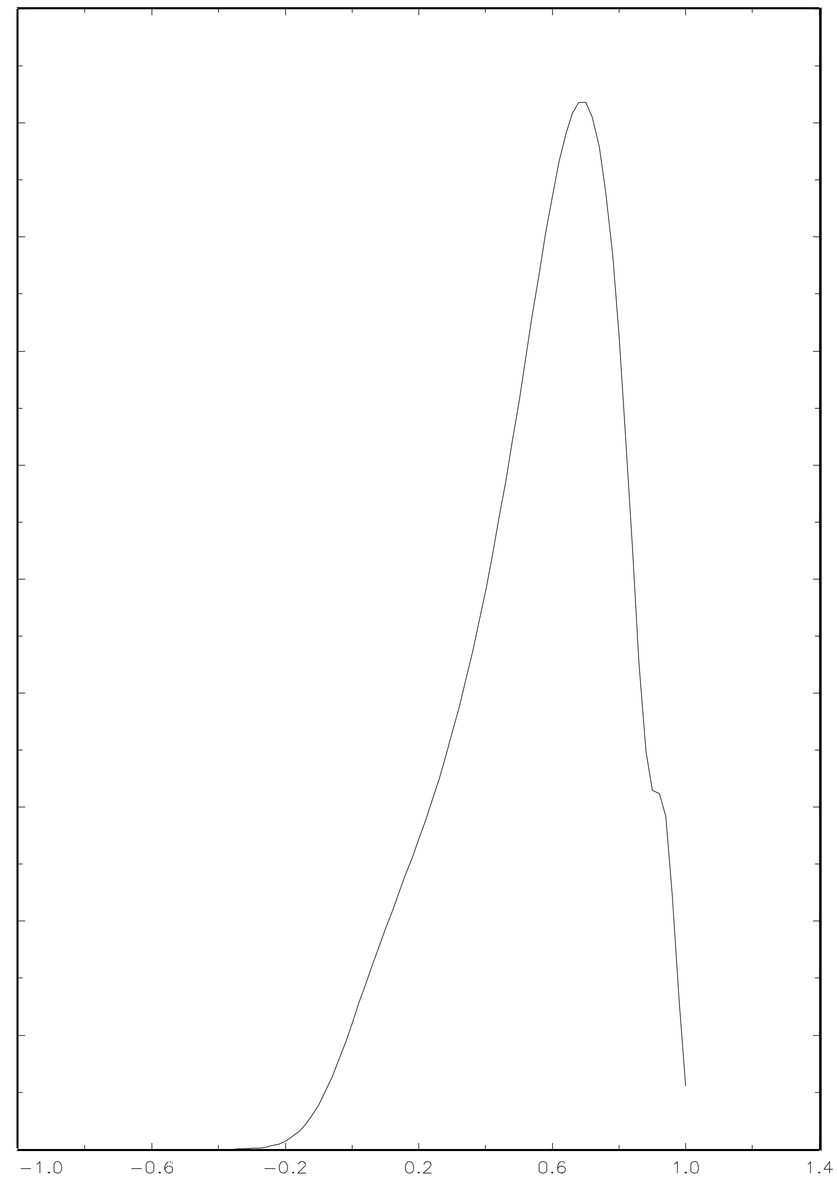

Figure 5a: Joint for Theta and Rho

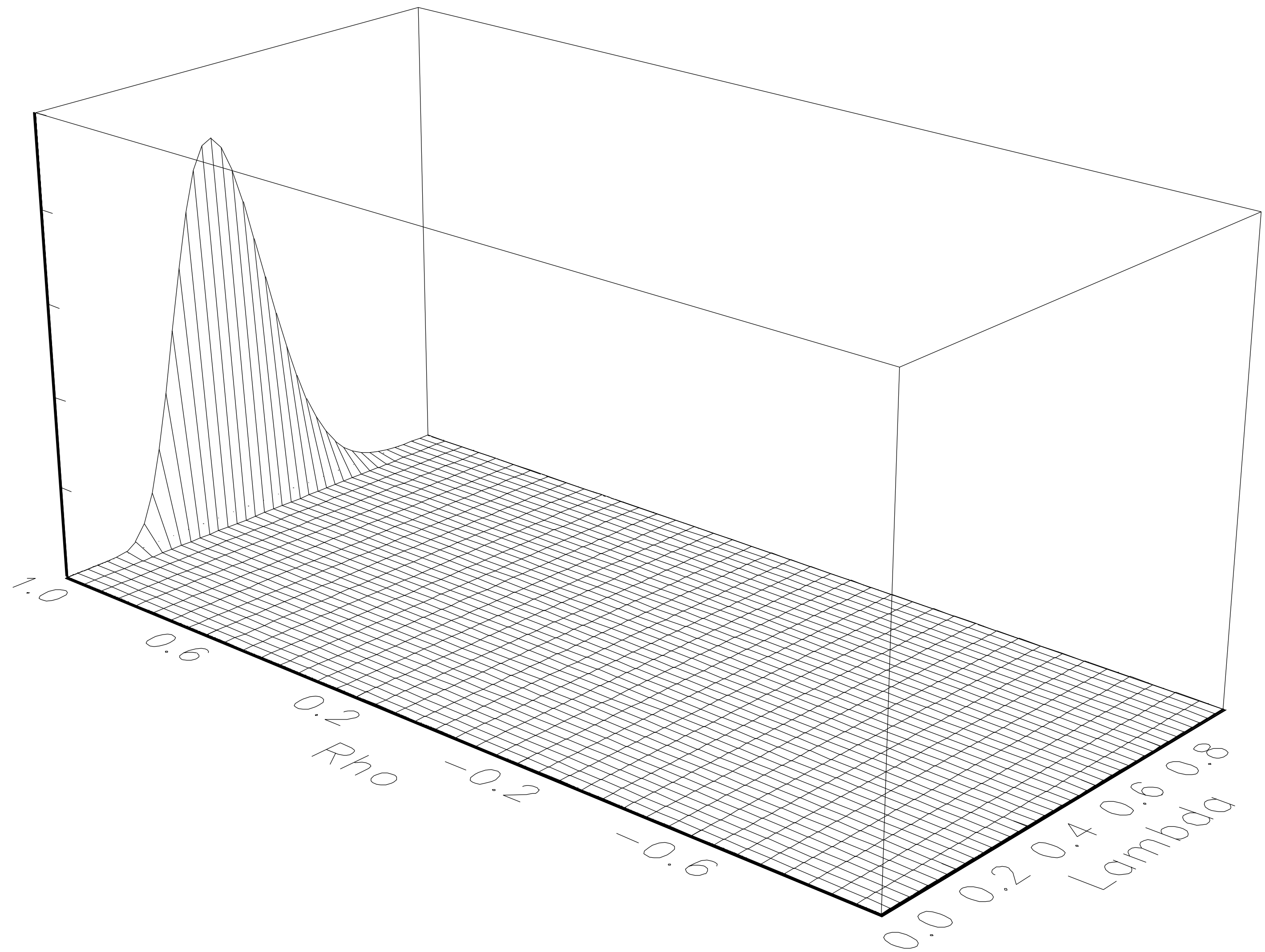

Figure 5b: Marginal for Theta

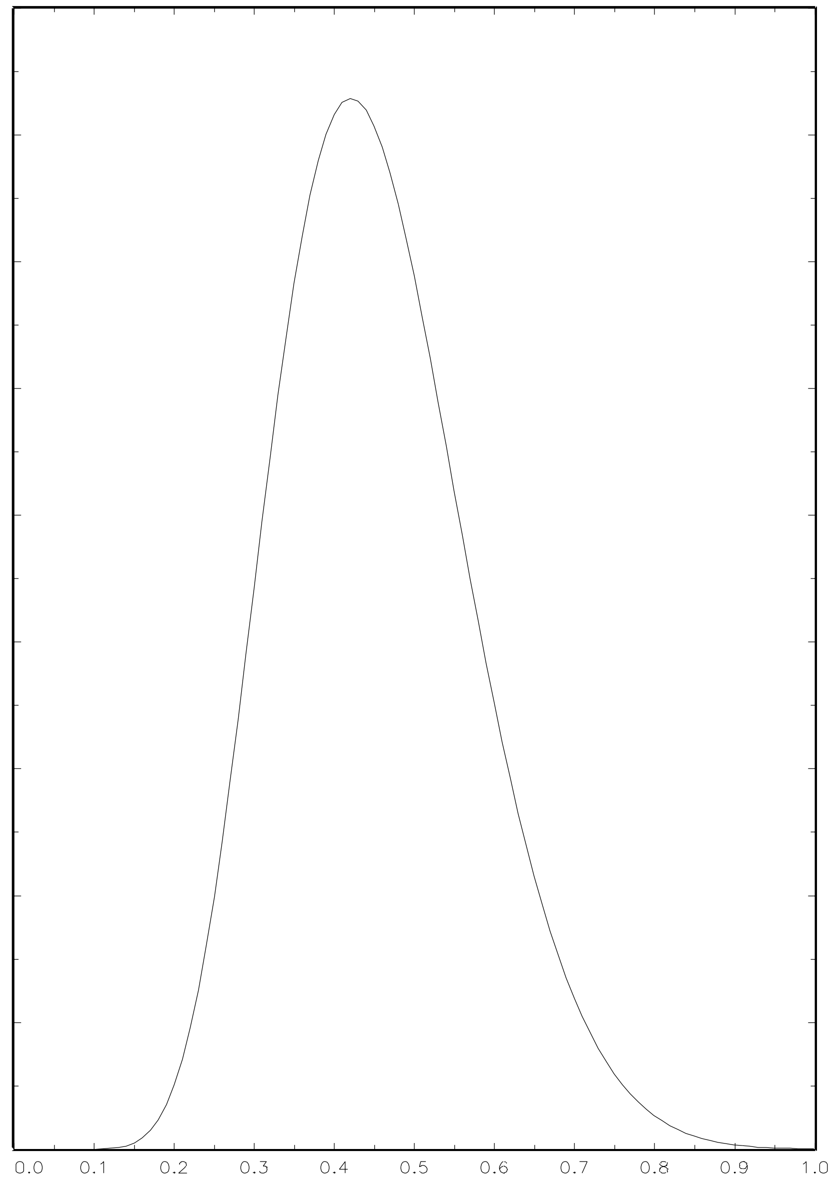

Figure 5c: Marginal for Rho

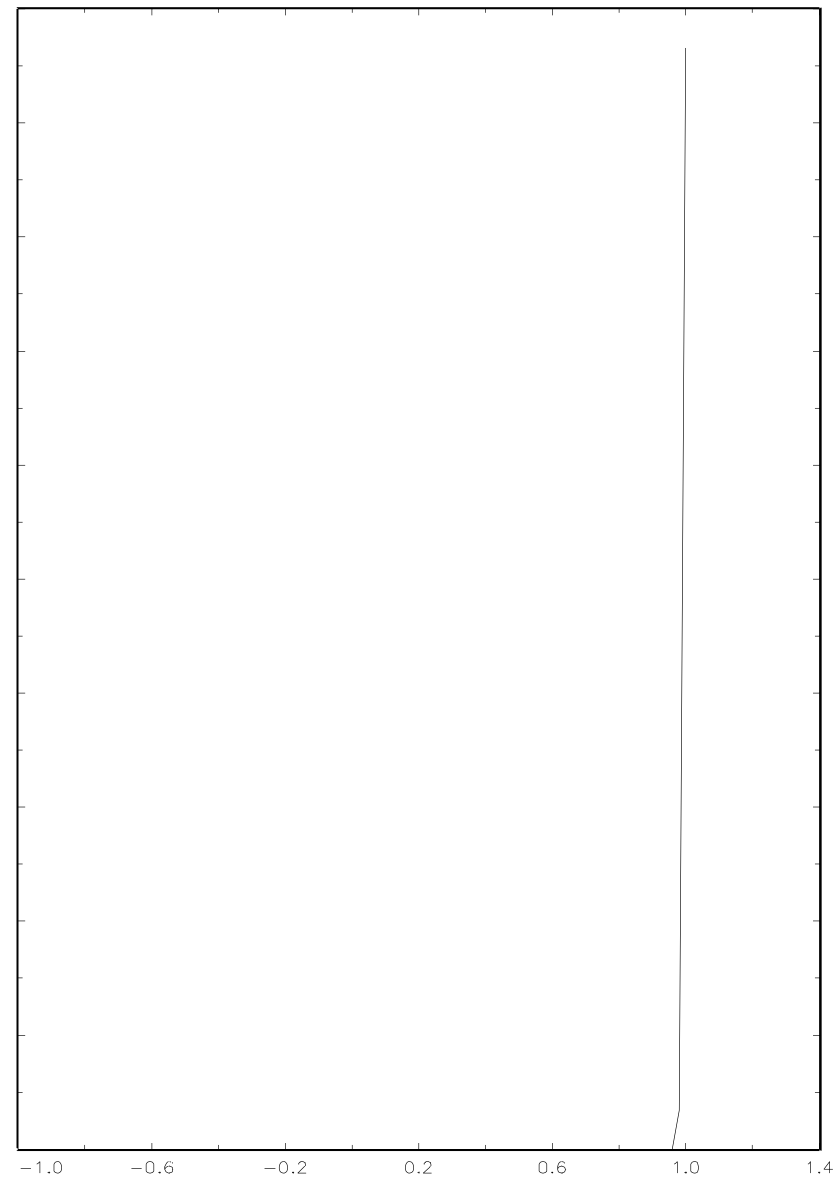

Figure 6a: Joint for Theta and Rho

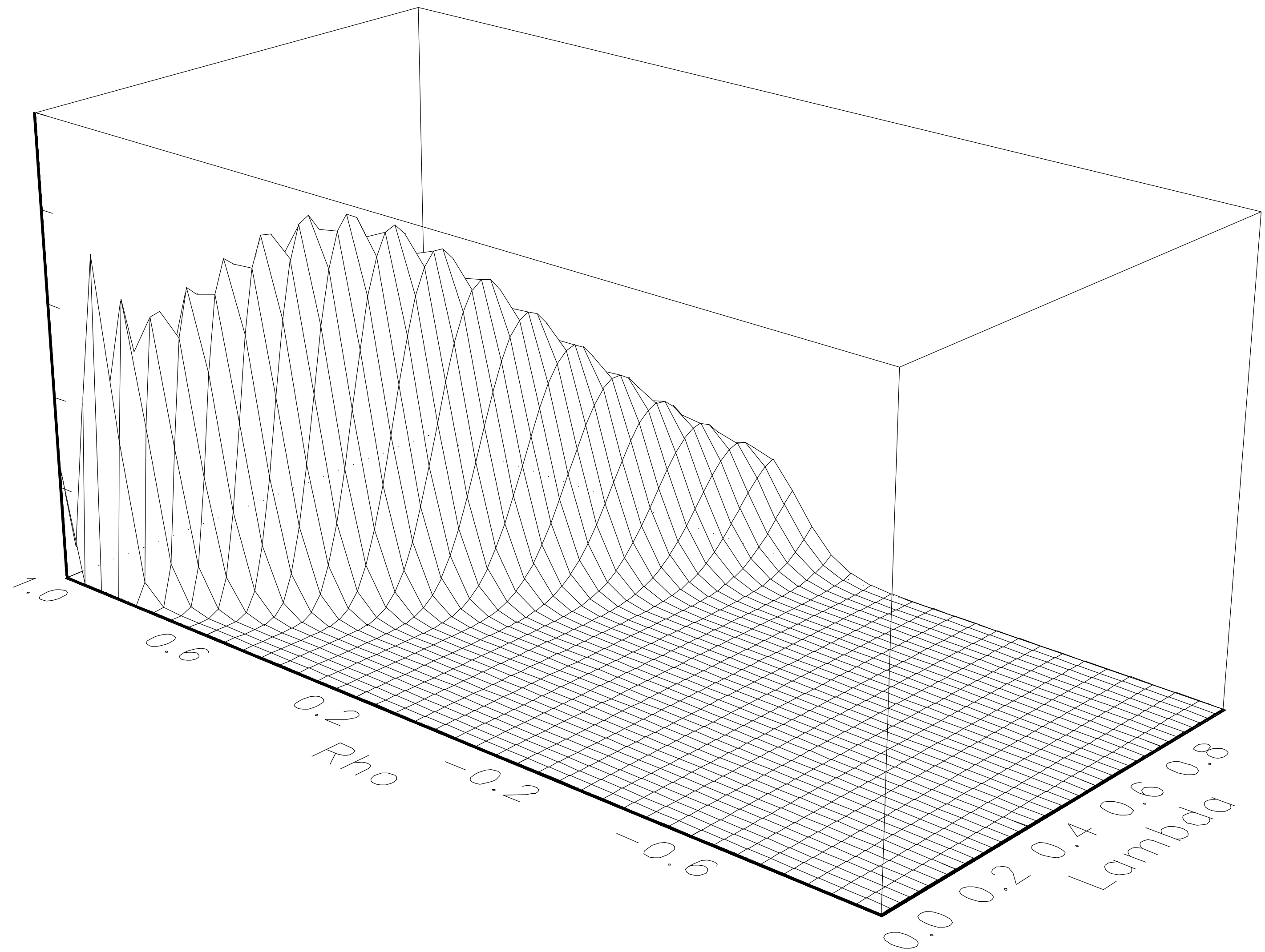

Figure 6b: Marginal for Theta

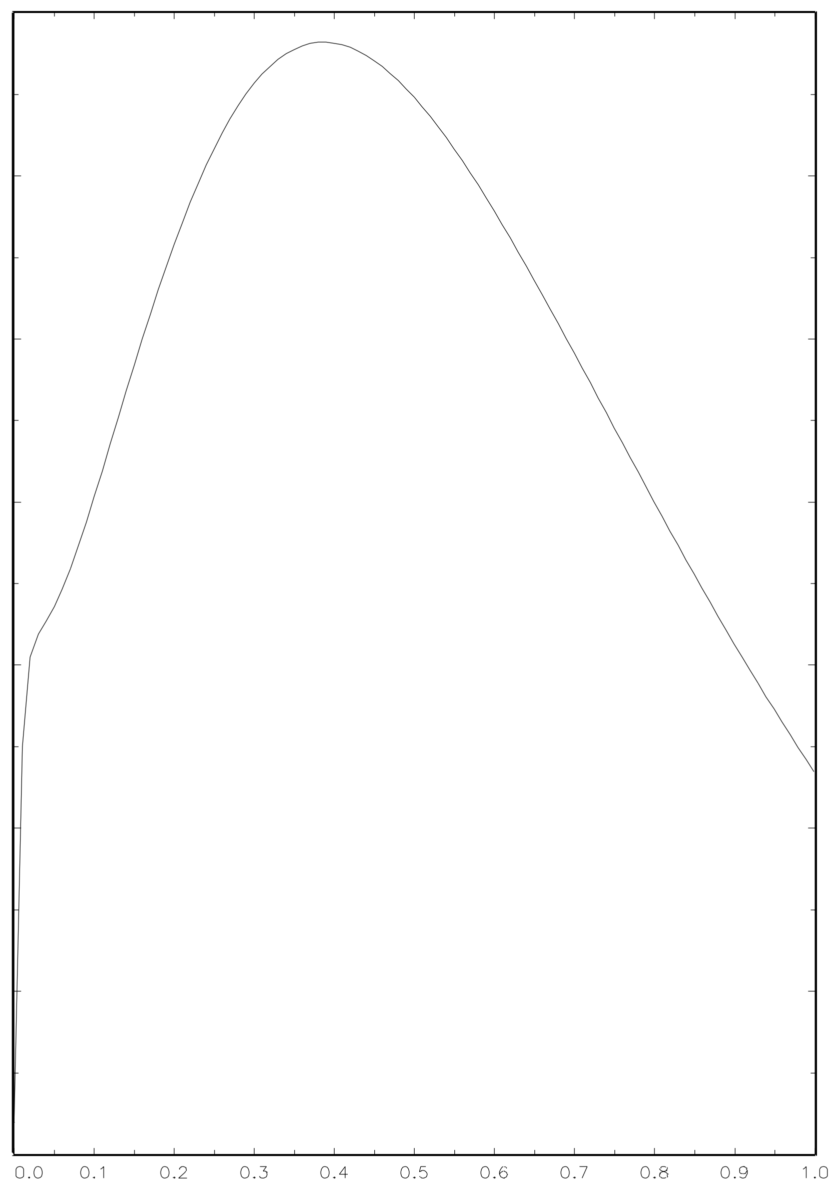

Figure 6c: Marginal for Rho

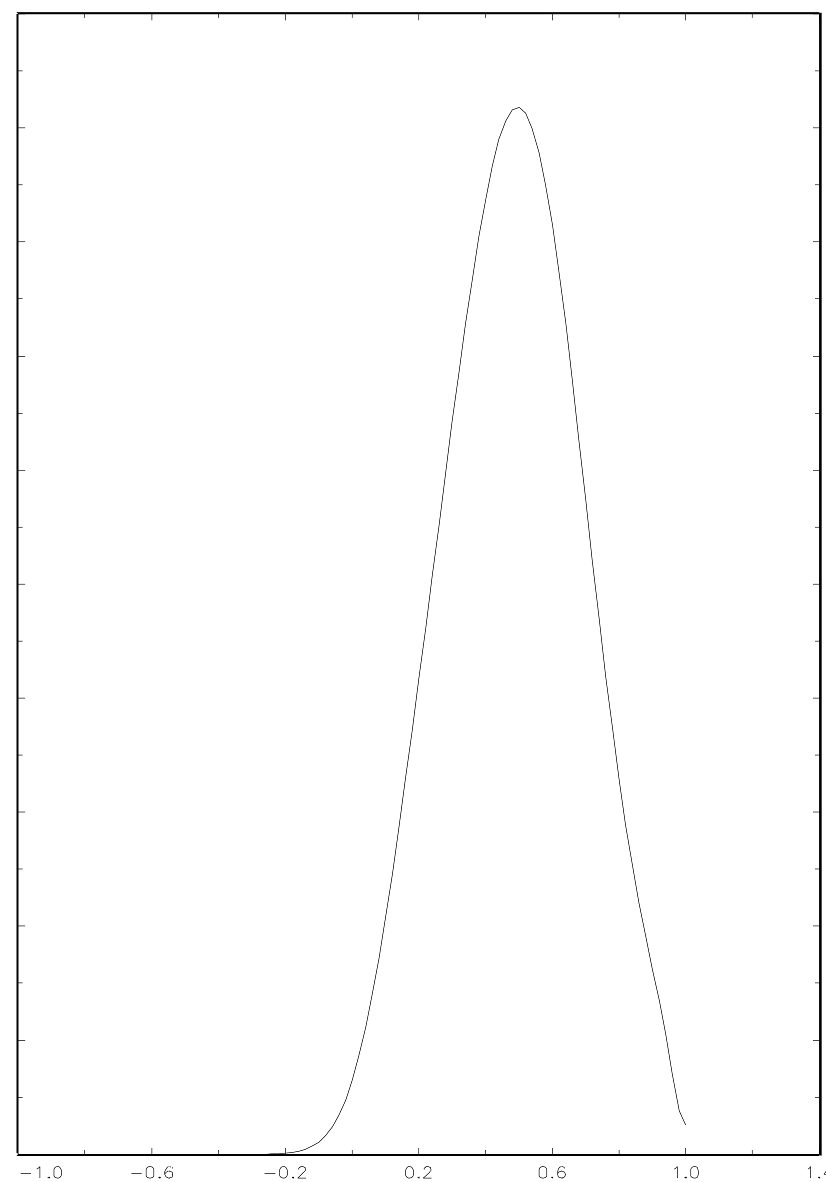

Figure 7a: Joint for Theta and Rho

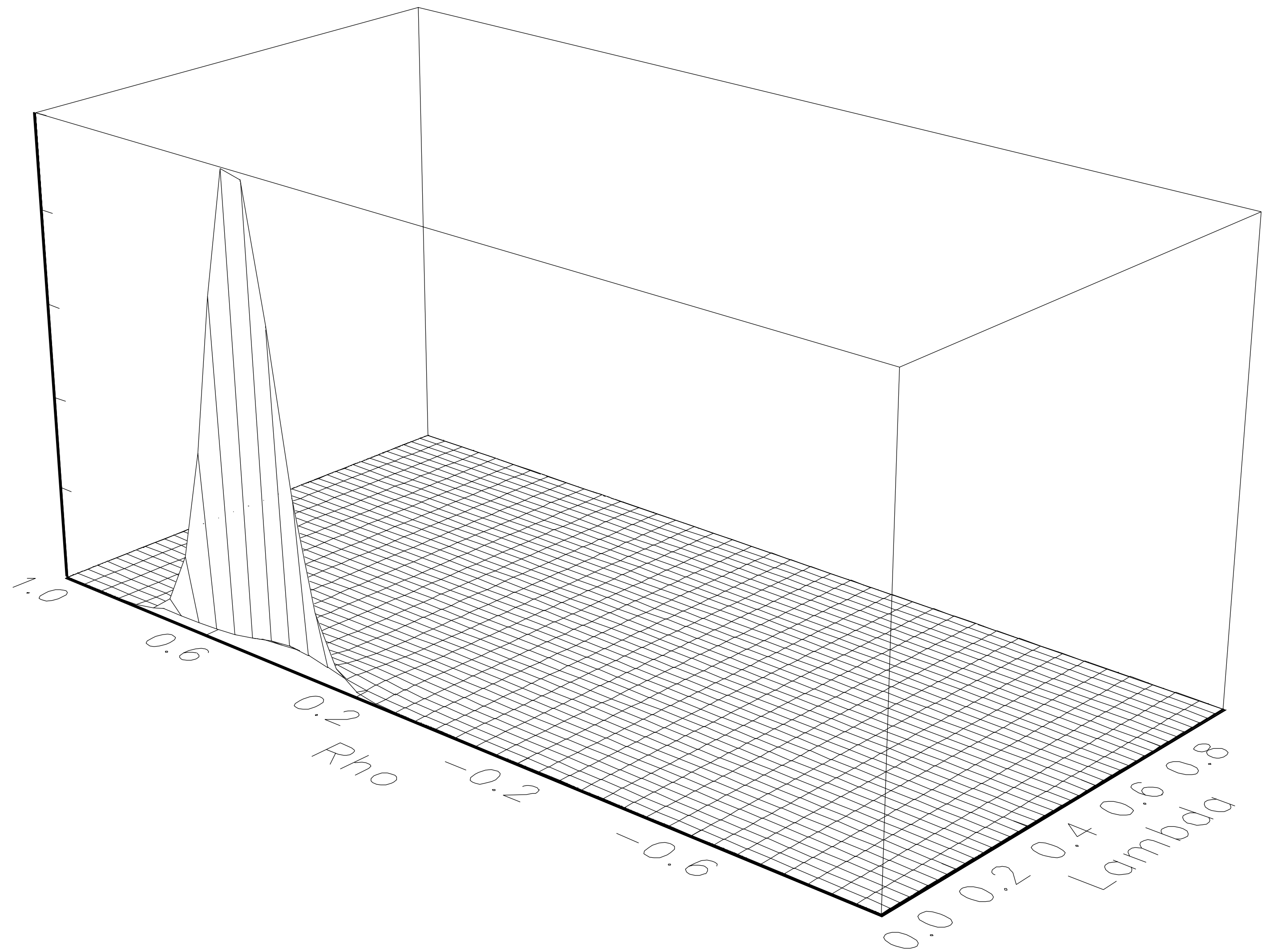

Figure 7b: Marginal for Theta

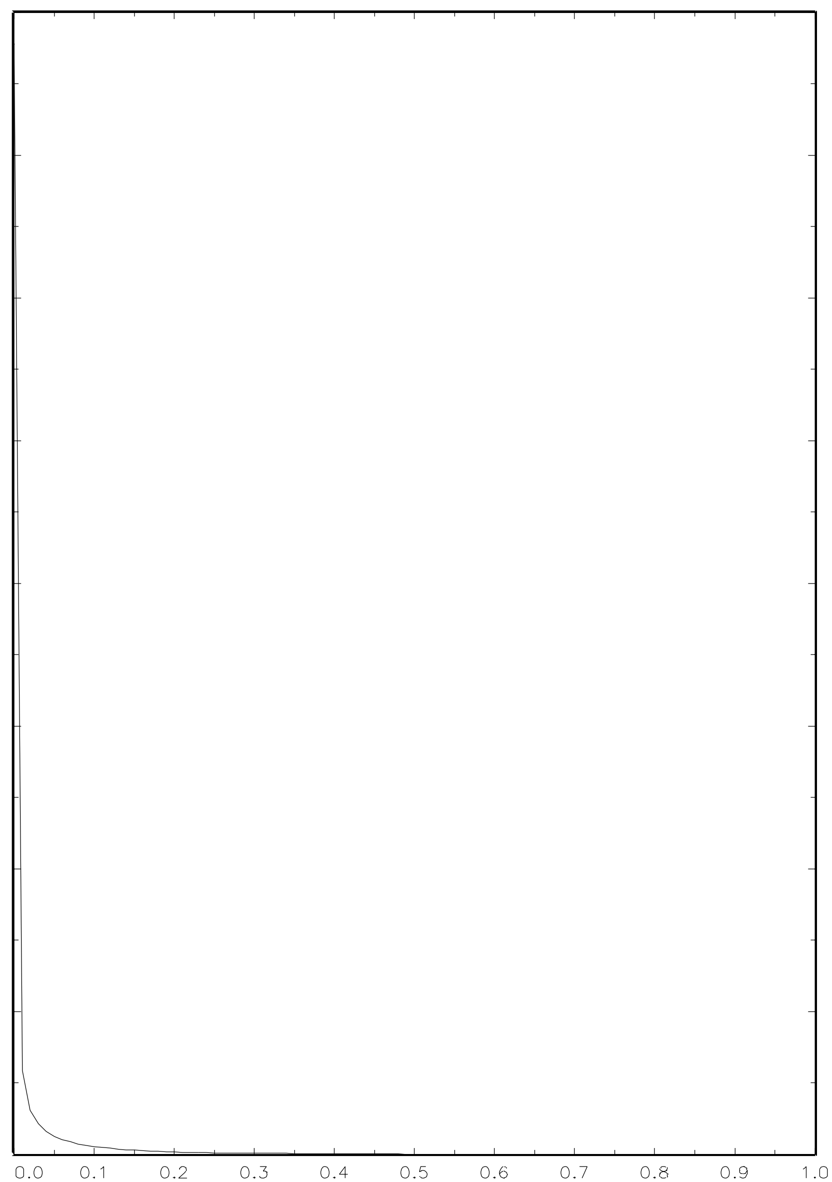

Figure 7c: Marginal for Rho