Variance-Covariance from a Metropolis Chain on a Curved, Singular Manifold*

A. Ronald Gallant†

First draft: April 10, 2021
This draft: August 22, 2022

Paper: www.aronaldg.org/papers/sdev.pdf
Code: www.aronaldg.org/webfiles/npb
Slides: www.aronaldg.org/papers/npbsdev.pdf

Forthcoming, *Journal of Econometrics*

*Address correspondence to A. Ronald Gallant, P.O. Box 659, Chapel Hill NC 27514, USA, phone 919-428-1130; email aronldg@gmail.com.

†Emeritus Professor of Economics, University of North Carolina; Emeritus Professor of Economics, Penn State University.

Abstract

We consider estimation of variance and covariance from a point cloud that are draws from a posterior distribution that lie on a curved, singular manifold. The motivating application is Bayesian inference regarding a likelihood subject to overidentified moment equations using MCMC (Markov Chain Monte Carlo). The MCMC draws lie on a singular manifold that typically is curved. Variance and covariance are Euclidean concepts. A curved, singular manifold is not typically a Euclidean space. We explore some suggestions on how to adapt a Euclidean concept to a non-Euclidean space then build on them to propose and illustrate appropriate methods.

Keywords and Phrases: Method of moments, Bayesian inference, Simultaneously valid credibility intervals, Point cloud, Curved, singular manifold.

JEL Classification: C11, C14, C15, C32, C36, C58

1 Introduction

There are many uses for scale measures. Among them are tuning an MCMC chain, determining the appropriate relative lengths of the sides of multivariate credibility rectangles, and adhering to the convention of reporting both location and scale in the presentation of statistical results.

If the MCMC chain[1] lies on a curved, singular manifold, the computation of scale becomes problematic because distance on a manifold is reckoned according to distance along a geodesic rather than along a strait line as in a Euclidean space. Conventional measures of scale such as[2] S^2 are based on straight line notions of distance and are centered at a point that does not necessarily lie on the manifold. This can lead to both understatement of scale and poor location of credibility rectangles relative to scale and location based on geodesics. This situation usually arises as follows.

A likelihood

$$f(y \,|\, x, \rho) = \prod_{t=1}^{n} f(y_t \,|\, x_{t-1}, \rho), \tag{1}$$

where y_t is a column vector and x_{t-1} is a matrix of exogenous and predetermined variables with a fixed number of rows, is available. The number of columns of x_{t-1} is either fixed, as in a VAR model or a cross-sectional application, or increasing with t, as in a VAR-GARCH model.[3] The y and x are objects that contain the observed y_t and x_{t-1}. In the applications envisioned here the likelihood is usually chosen to be a sieve with variable number of parameters thus making the Bayes estimator nonparametric, but this is not essential.

Estimation of the parameters in (1) is subject to moment conditions[4]

$$0 = q(\rho, \theta) = \frac{1}{n} \sum_{t=1}^{n} \int m(y, x_{t-1}, \rho, \theta) f(y \,|\, x_{t-1}, \rho)\, dy, \quad q \in \mathbb{R}^m \tag{2}$$

support conditions

$$h(\rho, \theta) > 0, \quad h \in \mathbb{R}^l \tag{3}$$

[1]The terms "MCMC chain", "MCMC draws", and "point cloud" are used interchangeably; "flat space" and "Euclidean space" are used interchangeably.

[2]For multivarite u_t, $S^2 = \frac{1}{n} \sum_{t=1}^{n} (u_t - \tilde{u})(u_t - \tilde{u})^\top$, where $\tilde{u} = \frac{1}{n} \sum_{t=1}^{n} u_t$.

[3]This is due to the recursive structure of GARCH variance which causes a VAR-GARCH model to be non-Markovian and to depend on the past up to the initial observation as most VAR-GARCH likelihoods are implemented in practice.

[4]One can integrate with respect to the distribution of x_{t-1} rather than the empirical distribution of x_{t-1} if it is available.

and a prior

$$\pi(\rho, \theta). \tag{4}$$

Letting[5] $\mathsf{x} = (\rho, \theta)$, the support of the posterior is the manifold

$$M = \left\{ \mathsf{x} \in \mathbb{R}^{d_a} : \; q_i(\mathsf{x}) = 0, \; i = 1, \ldots, m, \; h_j(\mathsf{x}) > 0, \; j = 1, \ldots, l \right\} \tag{5}$$

The parameters ρ are induced in $q(\rho, \theta)$ by the integration but they may also appear explicitly in $m(y_t, x_{t-1}, \rho, \theta)$. The parameters θ are those that appear only in the moment functions $m(y_t, x_{t-1}, \rho, \theta)$. We assume overidentification, i.e., that the dimension m of $q(\rho, \theta)$ is larger than the dimension of θ. Under this setup, the support of the posterior density is singular with respect to Lebesgue measure (Bornn, Shephard, and Solgi, 2018).

Our interest is in draws from algorithms that deliver exact solutions. That is, the draws $\{\mathsf{x}_i\}_{i=1}^N$ are in M to within reasonable precision for linear algebra on a machine. To our knowledge, there are three algorithms that can satisfy this requirement for the Bayesian inference problem defined by (1) through (4). Gallant (2022) generates draws $\{\mathsf{x}_i\}_{i=1}^N$ in M for the problem as stated by using the Surface Sampling Algorithm of Zappa, Holmes-Cerfon, and Goodman (2018). Bornn, Shephard, and Solgi require that (1) has discrete support, which makes (2) a sum involving probability weights and their corresponding support. Their paper contains numerous examples and an extensive review of literature related to this problem. Shin presumes that (1) is a mixture of specific parametric distributions with random weights drawn from a discrete distribution. The constraint (2) becomes a constraint on the discrete distribution of the random weights. His examples are from macro economics. If geodesics are explicitly available then the geodesic Monte Carlo method of Byrne and Girolami (2013) is available and the methods proposed here can be drastically simplified because the difficulty of only having a point cloud to work with is eliminated.

In what follows, requirements are as follows: The manifold M given by (5) must be nonempty and connected. Define

$$Q_{\mathsf{x}} = \left[\frac{\partial}{\partial \mathsf{x}} q_1(\mathsf{x}), \; \ldots, \; \frac{\partial}{\partial \mathsf{x}} q_m(\mathsf{x}) \right], \tag{6}$$

[5] In this paper, sans serif x and y are distinguished from italic x and y; the former referring to parameters and the later to data. This is to maintain compatibility with both econometric conventions and the numerical analysis conventions of Zappa, Holmes-Cerfon, and Goodman (2018) and Gallant (2022).

which is the transpose of the Jacobian of $q(\mathsf{x})$ and has dimension d_a by m. We assume that Q_{x} exists and has full rank for all $\mathsf{x} \in M$.

The functions $q(\mathsf{x})$ and $h(\mathsf{x})$ that determine M might arise from considerations other than the Bayesian inference problem defined by (1) through (4). Thus our results potentially have broader applicability. They apply to any random point cloud on M for which duplicate points either do not occur or can be easily identified. And to any M for which $q(\mathsf{x})$, which need not be of the form of (2), and Q_{x} can be computed. Conversely, if one has a convenient parametric representation of M such as $M = \{\mathsf{x} \in \mathbb{R}^{d_a} : \mathsf{x} = g(u), u \in \mathbb{R}^d, d < d_a\}$ and a convenient representation of geodesics on M such as $\gamma(t)$, then one might prefer the methods set forth in Pennec (2006).

Code, including a User's Guide, implementing methods introduced here for the SNP sieve $f(y_t \,|\, x_{t-1}, \rho)$ proposed by Gallant and Nychka (1987) as adapted to time series applications by Gallant and Tauchen (1989) is at `http://www.aronaldg.org/webfiles/npb`.

Parts of this paper borrow from Gallant (2022) so as to make this paper self contained.

2 Geodesics

On a manifold $M \subset \mathbb{R}^{d_a}$ of dimension $d < d_a$, distance is computed along geodesics. One computes distance by traversing a geodesic from a starting point s to an end point p and accumulating (infinitesimal increments of) a Hausdorff weight function $\delta_M(s, p)$ defined on M (Morgan, 2016).

As shown by Memoli and Sapiro (2001, Subsection 1.1), one can compute an approximate geodesic by putting an ϵ-offset on the manifold M to obtain a d_a-dimensional subset M_ϵ of $\mathbb{R}^{d_a}$ and applying the Fast Marching Algorithm of Sethian (1996). For a point cloud on M one can construct such an M_ϵ as the union of ϵ-balls centered at the points provided ϵ is large enough that M_ϵ is a connected set. Unfortunately, the Fast Marching Algorithm requires that M_ϵ be placed on equally spaced grid by interpolation. The demands of a grid on computer memory limit the applicability of the Fast Marching Algorithm to problems where d_a is less than about five. Regardless of dimension, the method described next is far more convenient for a point cloud.

Rather than an interpolated, equally spaced grid, one can let the point cloud determine

an unequally spaced grid and use Dijkstra's algorithm (Dijkstra, 1959) to compute geodesics. If M_ϵ is a connected set, then the MCMC draws may be viewed as nodes p_j of a graph $\mathcal{G}_\epsilon$ connected by edges $e_{j,j'}$ that have Euclidean length $\delta(p_j, p_{j'})$ and that stay within M_ϵ. In view of the fact that our point cloud is an MCMC chain and the contours of the density that the chain targets are not spheres, our ϵ-balls for determining $\mathcal{G}_\epsilon$ are rectangles with sides k equal to $\Delta \max\{|\mathsf{x}_{k,i} - \mathsf{x}_{k,i-1}| : \mathsf{x}_i \in \mathcal{D}\}$ where $\mathcal{D} = \{\mathsf{x}_i\}_{i=1}^N$ denotes the MCMC chain and $\mathsf{x}_{k,i}$ denotes the kth element of x_i. From a start s, Dijkstra's algorithm finds the shortest path that traverses edges to every node p_j. Computations are as follows.

The MCMC chain $\mathcal{D} = \{\mathsf{x}_i\}_{i=1}^N$ will contain duplicates due to rejections. They are easily detected because they must occur in succession. Nodes are the distinct points $\{p_j\}_{j=1}^{N^*}$ and $j(i)$ is the mapping from draw index i to node index j. Dijkstra's algorithm gives the distance $\delta(s, p_j)$ along edges from s to every node p_j and the path $(j_1^p, j_2^p, ..., j_k^p)$ that connects them, where j_1^p refers to starting node s and j_k^p to ending node p_j

One proceeds by choosing a Δ and constructing the graph $\mathcal{G}_\epsilon$. If $\mathcal{G}_\epsilon$ is not connected, Dijkstra's algorithm will return ∞ for the distance from s to an isolated node. One can find the smallest admissible Δ by finding a Δ with isolated nodes then increasing Δ until Dijkstra's algorithm does not find isolated nodes. One can check that edges lie within M_ϵ by seeing if the midpoint of every edge is in M_ϵ to within a reasonable tolerance.

3 Estimating Scale

The notions of mean, variance, and covariance are flat space concepts, i.e., Euclidean space concepts, and it is not obvious how to extend them to a curved, nonlinear manifold.

There seems to be general agreement on how to define a mean over a curved, nonlinear manifold and estimate it from a sample (Wikipedia, 2021). The estimate of the intrinsic mean, $\bar{\mathsf{x}}$, is the start s that minimizes $\frac{1}{N}\sum_{i=1}^N \delta^2(s, p_{j(i)})$. This is an example of a Frechet mean on a metric space (Pennec, 1999, Sec. 2.2). The term $\frac{1}{N}\sum_{i=1}^N \delta^2(s, p_{j(i)})$ is an example of a Frechet variance. It is a sort of total variation. Because it does not provide variances and correlations coordinate by coordinate it is not adequate for the applications we envisage here.

Computing the intrinsic mean is an order N^2 computation and is quite time consuming.

One way to reduce run times is to reduce the length of the chain by retaining, say, every tenth element of the chain. Chains are often highly serially correlated in which case reducing the length of the chain in this fashion does not lose much information. For large Δ the number of edges in the graph $\mathcal{G}_\epsilon$ approaches $\frac{1}{2}N(N-1)$. One may have to reduce the length of the chain anyway to keep file sizes within reason and the run times of Dijkstra's algorithm within reason.

Another way to reduce run times is to search only among likely candidates for the mean. One can, say, divide the chain into ten segments and unrestrictedly search for the intrinsic mean in the first segment. Then search only among, say, the thousand nodes closest to the mean found in the first segment in the combined first and second segments. Continue so on until one is searching among, say, the closest one hundred in the full chain. Fortunately the estimators of scale that we discuss below do not appear sensitive to slight errors in finding the intrinsic mean.

A first approach to estimating variances and covariances is to ignore the fact that the MCMC chain lies on a manifold. One computes a mean by averaging over the chain: $\tilde{\mathsf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathsf{x}_i$. This is the extrinsic mean. One computes sample variances and covariances by averaging squares and cross products of deviations from the extrinsic mean $\tilde{\mathsf{x}}$ over the chain:

$$V_{EC} = S^2 = \frac{1}{N}\sum_{i=1}^{N}(\mathsf{x}_i - \tilde{\mathsf{x}})(\mathsf{x}_i - \tilde{\mathsf{x}})^\top. \tag{7}$$

One could, instead, perform the same computation but center at the intrinsic mean $\bar{\mathsf{x}}$:

$$V_{IC} = \frac{1}{N}\sum_{i=1}^{N}(\mathsf{x}_i - \bar{\mathsf{x}})(\mathsf{x}_i - \bar{\mathsf{x}})^\top. \tag{8}$$

The diagonals of V_{EC} are smaller than those of V_{IC}.

One can make the following argument in support of extrinsic variance and covariance as opposed to the approaches described below. The manifold M can be viewed as a sample space and $\{\mathsf{x}_i\}_{i=1}^N$ as a sample from the distribution P that the chain targets. The function that maps the j,k elements of $\mathsf{x} \in M$ to the two dimensional reals,

$$f_{j,k}(\mathsf{x}) \,:\, \mathsf{x} \mapsto \mathbb{R}^2 \tag{9}$$

is a random variable on the probability space $(M, \mathcal{B}, P)$, where $\mathcal{B}$ denotes the Borel subsets of M. One can argue that the two variances and the covariance of this random variable are

the objects of interest. If one adopts this view, then V_{EC} is the correct computation. Using, V_{IC} instead of V_{EC} makes a slight accommodation to the origin of the chain by centering at a point $\bar{\mathsf{x}}$ that is in M.

A third approach is essentially the same as an extrinsic computation but one increases each coordinate of a point p_j by its geodesic distance in that direction. Specifically, for the path[6] $(j_1^p, j_2^p, ..., j_k^p)$ that connects $\bar{\mathsf{x}}$ to p_j, where j_1^p indexes node $\bar{\mathsf{x}}$ and j_k^p indexes node p_j, put[7]

$$D_j = \text{diag}[\text{sgn}(p_j - \bar{\mathsf{x}})] \sum_{\ell=2}^{k} |p_{j_\ell^p} - p_{j_{\ell-1}^p}| \quad D_j \in \mathbb{R}^{d_a} \tag{10}$$

where the signum and absolute value functions are applied elementwise to $p_j - \bar{\mathsf{x}}$ and $p_{j_\ell^p} - p_{j_{\ell-1}^p}$, respectively. The estimated variance-covariance matrix is

$$V_{ME} = \frac{1}{N} \sum_{i=1}^{N} D_{j(i)} D_{j(i)}^\top. \tag{11}$$

We will call the third approach modified extrinsic variance-covariance.

The diagonals (variances) of V_{ME} are larger than the diagonals of V_{IC}. The larger is Δ the smaller are the diagonals of V_{ME}. If one sets Δ so large that all nodes are connected, then V_{ME} is equal to V_{IC} to within rounding errors caused by thrusting Dijkstra's algorithm into the middle of the computation.

The modified extrinsic approach makes intuitive sense. Consider three dimensional space and let the variance under consideration be that for altitude. If the altitude of the end point p_j and the intrinsic mean $\bar{\mathsf{x}}$ are both the same then the contribution of p_J to the diagonal element of V_{IC} corresponding to altitude is zero. If there is a ridge on the manifold between $\bar{\mathsf{x}}$ and p_j then the contribution to the modified extrinsic variance V_{ME} will be the square of the sum of the two changes in altitude that it takes to ascend and descend the ridge. This seems to be a reasonable representation of the geometry of the situation.

A defect of the above approaches is that the manifold M is singular with respect to Lebesgue measure but the matrices V_{EC}, V_{IC}, and V_{ME} will not necessarily be singular. The next approach respects singularity.

[6]This path is an output of Dijkstra's algorithm and k is part of that output.

[7]This expression corrects an error in equation (15) of Gallant (2022).

The fourth approach uses notions from Riemannian geometry; a good reference is Pennec (2006). The idea is to represent the manifold as a flat space called a chart and then compute variances and covariances in the usual way on the chart. The chart is a Euclidean space if the inner product under the coordinate system chosen to represent points on the chart is Euclidean, as we assume. For instance, if the manifold were the surface of the earth with elevations disregarded, the chart would be a two dimensional world map and the MCMC chain would map to points on this world map.

The flat space one uses as a chart with the Riemannian approach is the plane $T_{\bar{\mathsf{x}}}M$ tangent to the manifold M at the intrinsic mean $\bar{\mathsf{x}}$. A point x_i from the MCMC chain $\{\mathsf{x}_i\}_{i=1}^N$ on M is plotted on this chart as follows. For each geodesic $\gamma(t)$ with $\gamma(0) = \bar{\mathsf{x}}$, the tangent vector $\frac{d}{dt}\gamma(0)$ is in $T_{\mathsf{x}}M$. Let $\gamma_i(t)$ be the geodesic connecting $\bar{\mathsf{x}}$ to the point x_i for which distance $\delta_i = \delta(\bar{\mathsf{x}}, \mathsf{x}_i)$ along the geodesic is smallest and let $\hat{v}_i$ be the tangent vector $\hat{v}_i = \frac{d}{dt}\gamma_i(0)$. The marker $\hat{z}_i$ corresponding to x_i is placed on the chart $T_{\bar{\mathsf{x}}}M$ at $\hat{z}_i = \delta_i \frac{\hat{v}_i}{\|\hat{v}_i\|}$.

There are some technical problems with this approach, the most important of which is that the geodesic with smallest distance $\delta(\bar{\mathsf{x}}, \mathsf{x}_i)$ may not be unique. These can be addressed in ways that need not concern us because the method is not feasible when all one has available is an MCMC chain on M. But we can borrow the ideas of using the tangent plane as a chart, mapping x_i from the chain to $\hat{z}_i$ on the chart by means of lines emanating from $\bar{\mathsf{x}}$, and making the distance of $\hat{z}_i$ from $\bar{\mathsf{x}}$ along the line in the chart the same as the distance $\delta(\bar{\mathsf{x}}, \mathsf{x}_i)$ from $\bar{\mathsf{x}}$ on the manifold.

To modify the Riemann approach to be applicable to a chain on a manifold we need to explicitly define a basis on $T_{\bar{\mathsf{x}}}M$: Put $A = [Q_{\bar{\mathsf{x}}} \,|\, 0\,]$, where $Q_{\bar{\mathsf{x}}}$ is given by (6). A is a square matrix of dimension d_a by d_a whose last $d = d_a - m$ columns are filled with zeros. Apply the singular value decomposition algorithm (Businger and Golub, 1969) to obtain $A = USV^\top$; U will be orthogonal and S diagonal with the first m diagonal entries positive and the remainder zero. If S is not such, $Q_{\bar{\mathsf{x}}}$ does not have full rank, which violates the regularity conditions set forth in Section 1. Partition U as $\left[T_{\bar{\mathsf{x}}}^{\perp} \,|\, T_{\bar{\mathsf{x}}}\right]$, where $T_{\bar{\mathsf{x}}}^{\perp}$ has m columns and $T_{\bar{\mathsf{x}}}$ has d columns.

A point x_i from from the MCMC chain $\{\mathsf{x}_i\}_{i=1}^N$ on M is plotted on the chart $T_{\bar{\mathsf{x}}}M$ as follows. Put $v_i = T_{\bar{\mathsf{x}}}T_{\bar{\mathsf{x}}}^\top(\mathsf{x}_i - \bar{\mathsf{x}})$; v_i is the orthogonal projection of $\mathsf{x}_i - \bar{\mathsf{x}}$ onto $T_{\bar{\mathsf{x}}}M$. The

marker z_i corresponding to x_i is placed on the chart $T_{\bar{\mathsf{x}}}M$ at $z_i = \delta_i \frac{v_i}{\|v_i\|}$. Define the modified Riemann variance-covariance matrix as

$$V_{MR} = \frac{1}{N} \sum_{i=1}^{N} z_i z_i^\top \tag{12}$$

Note that V_{MR} is d_a by d_a with rank d.

To compare the modified Riemann approach to the Riemann approach we return to the world map analogy with radius of the earth equal to one and elevations disregarded. Both charts would be circles of radius π. Suppose that all the MCMC draws $\{\mathsf{x}_i\}_{i=1}^N$ were on the two geodesics pointing north and northeast from $\bar{\mathsf{x}}$. The Riemann chart and the modified Riemann chart would look the same with the markers $\{z_i\}_{i=1}^N$ plotted on two straight lines emanating from $\bar{\mathsf{x}}$ that diverge. The antipodal $\bar{\mathsf{x}}_{ap}$ point cannot be plotted on either chart. In the Riemann approach this is because both geodesics connect to $\bar{\mathsf{x}}_{ap}$ with the same distance. In the modified Riemann approach it is because $\bar{\mathsf{x}}_{ap}$ plots atop $\bar{\mathsf{x}}$. One consequence is that the chart is an open set that does not contain a marker for $\bar{\mathsf{x}}_{ap}$. One has to presume that $\bar{\mathsf{x}}_{ap}$ occurs with probability zero.

Now put a symmetric mountain (a cone) due north of $\bar{\mathsf{x}}$ with a base large enough that the mountain intersects the northeast set of draws. The modified Riemann chart would look the same other than no longer being a circle due to a bulge at the terminus of the north and northeast straight lines emanating from $\bar{\mathsf{x}}$ and altered spacing of the markers when the mountain is encountered. The Riemann chart would look different. All north points on the far side of the mountain would have to be deleted because two different geodesics originating at $\bar{\mathsf{x}}$ connect to them at the same distance. These points would have to be presumed to occur with probability zero. Markers for the northeast points would fan out because the points no longer lie on a single geodesic.

3.1 Illustration of Scale Measures

Figure 1 illustrates the computations graphically. The panels (a), (b), (c), and (d) illustrate the computation of V_{EC}, V_{IC}, V_{ME}, and V_{MR}, respectively. Details are in the figure legend.

Figure 1 about here.

To illustrate the population quantities corresponding to V_{EC}, V_{IC}, V_{ME}, and V_{MR} consider the distribution on the half sphere

$$\begin{aligned}
(u, v) &\sim N_2(0, \Sigma) && (13)\\
\Sigma &= \frac{1}{4}\begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix} && (14)\\
r &= \sqrt{u^2 + v^2} && (15)\\
(x, y, z) &= \left(u, v, \sqrt{1 - r^2}\right) \text{ if } r < 1 \text{ else discard} && (16)
\end{aligned}$$

The results below are obtained from a simulation of (16) of size $N = 200000$. Due to discards when $r \geq 1$, the effective sample size was $n = 198373$.

The intrinsic mean, $\bar{\mathsf{x}}$ is $(0, 0, 1)$. The extrinsic mean, $\tilde{\mathsf{x}}$, is $(0, 0, 0.91656)$.

Computation of V_{EC}, V_{IC}, and V_{MR} from their definitions is straightforward. To compute D for V_{ME} the expression is

$$\begin{aligned}
D &= [x, y, z - 1 - |r\alpha - 2\sin(\alpha/2)|] && (17)\\
\alpha &= \arcsin(r) && (18)
\end{aligned}$$

The term $r\alpha - 2\sin(\alpha/2)$ is the length of the arc from $(0, 0, 1)$ to $(x, y, 1 - z)$ minus the length of the chord from $(0, 0, 1)$ to $(x, y, z - 1)$. This is the extra length of traveled in the z direction due to the curvature of the half sphere; note that z is always less than 1.

Denote the correlation matrices corresponding to V_{EC}, V_{IC}, V_{ME}, and V_{MR} by C_{EC}, C_{IC}, C_{ME}, and C_{MR}. Population values of these matrices are presented in Table 1.

For this example, the z axis is perpendicular to the tangent plane at the intrinsic mean. Due to this, the fact that V_{MR} is singular is obvious at sight in Table 1.

This highlights a deficiency of the modified Riemann variance-covariance matrix V_{MR} that is caused by trying to express standard deviations in the Euclidean coordinate system on $\mathbb{R}^{d_a}$ rather than the coordinate system $T_{\bar{\mathsf{x}}}$ on the chart: The entire point cloud must be close to the tangent plane, i.e., to the chart. Otherwise, scale can be understated as seen for z in the preceding example. A partial remedy is discussed the next section.

Table 1 about here.

3.2 Credibility Regions

Because the point cloud under consideration $\{\mathsf{x}_i\}_{i=1}^N$ is presumed to be a (possibly correlated) sample from a posterior distribution, it can be used to estimate the posterior probability of a set R. The set R may either be a proper subset of M or be a subset of $\mathbb{R}^{d_a}$. An estimated posterior probability can be computed as

$$P(R \,|\, x, y) = \frac{1}{N} \sum_{i=1}^{N} I_R(\mathsf{x}_i) \tag{19}$$

One would prefer to work with regions R that are proper subsets of M. This is to some extent possible but its practical use is severely inhibited by the difficulty of either visualizing or describing a curved, singular manifold in $\mathbb{R}^{d_a}$ for $d_a > 3$.

There is one extrinsic region of special interest

$$R_\tau = \times_{i=1}^{d_a} [\,\bar{\mathsf{x}}_i - \tau \,\mathrm{sdev}(\mathsf{x}_i),\ \bar{\mathsf{x}}_i + \tau \,\mathrm{sdev}(\mathsf{x}_i)], \tag{20}$$

where the $\mathrm{sdev}(\mathsf{x}_i)$ are the square roots of the diagonal elements of whichever of V_{EM}, V_{IM}, V_{ME}, or V_{MR} one prefers. Choosing τ such that $P(R_\tau \,|\, x, y) = 1 - \alpha$ gives a set of simultaneously valid $(1 - \alpha) \times 100\%$ credibility intervals.

Obviously one can base credibility regions on quantiles or other sorts of sets should one prefer.

Returning to consideration of the Riemann approach, the draws $\{\mathsf{x}_i\}_{i=1}^N$ in $\mathbb{R}^{d_a}$ that are expressed in terms of elementary basis vectors are in one-to-one correspondence with the points $\{z_i\}_{i=1}^N$ on the chart $T_{\bar{\mathsf{x}}}M$ that are expressed in terms of the basis vectors $T_{\bar{\mathsf{x}}}$. Therefore, a credibility region constructed on the chart, which is a flat space and therefore amenable to standard statistical analysis, can be mapped to the manifold M in $\mathbb{R}^{d_a}$. One can do this discretely by mapping the points z_i in the credibility region on the chart to the corresponding points x_i on the manifold M. Or, as an aid to visualization, one can regress the points x_i on the points z_i using, say, a multivariate polynomial and thereby represent a continuous region on the chart as a continuous region on the manifold. An example is Figure 2 which plots a slice of a 95% credibility rectangle on the chart to a coordinate of M. As a practical matter, for $d_a > 3$, one will probably have to restrict attention to regions of the form (20) that can be described in tabular form.

Figure 2 about here.

3.3 Most Likely Point

This subsection uses ideas from a referee to whom credit is due.

A measure of location that is on the surface of the manifold is the posterior mode. It is trivially easy to compute from an MCMC chain because the posterior is known to within the normalization constant at the accept/reject step of Metropolis algorithms; its computation is, therefore, essentially costless.

We followed tradition in this literature and used the intrinsic mean as the measure of location; see Pennec (1999, 2006) and the references therein. Early experience suggested that the posterior mode is often far enough away from the intrinsic mean that it does not make a good start for computation of the intrinsic mean. After finding an effective start strategy, little thought was given to the posterior mode. The referee suggests that this is evidence of skewness. Another possibility is that, while close in Euclidian distance, the posterior mode can be far enough away in geodesic distance to be a poor start due to the shape of the manifold. See, e.g., Figure 2, wherein the manifold folds tightly back on itself.

The argument advanced by the referee in favor of the posterior mode is in the context of credibility regions. The referee recommends obtaining a credibility region using

$$R_{1-\alpha} = \{\mathsf{x}_i \mid \log\, p(\mathsf{x}_i \mid x, y) > p_{1-\alpha}\} \tag{21}$$

where $p(\mathsf{x}_i|x, y)$ is the posterior density and $\{\mathsf{x}_i\}_{i=1}^N$ is the point cloud. We shall compare the referee's recommendation to

$$R^*_{1-\alpha} = \{\mathsf{x}_i \mid (x_i - \bar{\mathsf{x}})^\top V_{IC}^{-1}(x_i - \bar{\mathsf{x}}) < p^*_{1-\alpha}\}. \tag{22}$$

Both $R_{1-\alpha}$ and $R^*_{1-\alpha}$ were computed with $\alpha = 0.05$ for the demand and supply point cloud of Subsection 4.1. The surface area of the credibility region $R_{1-\alpha}$ may be computed by importance sampling, viz.

$$A = \sum_{\mathsf{x}\in R_{1-\alpha}} 1/\exp[\mathtt{const} + \log p(\mathsf{x} \mid x, y)] \tag{23}$$

and the same for A^*, the surface area of $R^*_{1-\alpha}$, with the same `const`.[8] The ratio of these areas is $A^*/A = 2.77$ thus bearing out the referee's claim that the region $R_{1-\alpha}$ is preferred.

In absolute value, the largest Fisher-Pearson coefficients of skewness for the elements $\mathtt{x}_i$ of $\mathtt{x}$ in the point clouds of the examples of Sections 4.1, 4.2, and 4.3 are 0.30675, 0.79980, and 0.11948, respectively. This is evidence against skewness except that marginalization discards surface information such as shown in Figure 2.[9] If one replaces the intrinsic mean, $\bar{\mathtt{x}}$, by the posterior mode in (22), the ratio of areas is 2.82, which is a small increase from 2.77. This is stronger evidence against skewness.

4 Examples

The first two examples are taken from Gallant(2022). Some of the text below is edited excerpts form that source.

4.1 A Simple Demand and Supply Example

Consider a simulation of the demand and supply system

$$
\begin{aligned}
x_t &= (\sigma_x + \rho_x x_{t-1}) z_{1,t} && (24)\\
\log q_{d,t} &= a_1 + a_2 \log p_t + \sigma_d z_{2,t} && (25)\\
\log q_{s,t} &= b_1 + b_2 \log p_t + x_t + \sigma_s z_{3,t} && (26)
\end{aligned}
$$

with solution $(\log p_t, \log q_t)$ under $q_t = q_{d,t} = q_{s,t}$, where $\sigma_x = 3$, $\rho_x = 0.2$, $a_1 = 12$, $a_2 = -2$, $b_1 = 3$, $b_2 = 4$, $\sigma_d = \sigma_s = 0.1$, $z_{i,t}$ standard normal, and sample size $n = 500$. Note that the supply shifter x_t is heteroscedastic with variance dependent on x_{t-1} whence the same for price p_t and quantity q_t. The data are $y_t = (\log p_t, \log q_t, x_t)$ for $t = 1, 2, \ldots, n$.

The likelihood used for estimation is normal with heteroscedastic errors that depend on past values of y_t:

$$
\begin{aligned}
y_t &\sim n_3(y_t \,|\, \mu, \Sigma_{t-1}) && (27)\\
\Sigma_{t-1} &= RR' + P(y_{t-1} - \mu)(y_{t-1} - \mu)'P', && (28)
\end{aligned}
$$

[8] The `const` can be chosen to enhance numerical stability because it cancels in the ratio A^*/A.

[9] See the discussion of Equation (9).

where R is upper triangular, and P is diagonal. Thus,

$$\rho = (\mu_1, \mu_2, \mu_3, R_{1,1}, R_{1,2}, R_{2,2}, R_{1,3}, R_{2,3}, R_{3,3}, P_{1,1}, P_{2,2}, P_{3,3}) \in \mathbb{R}^{12}.$$

A set of moment conditions for estimation of the demand equation (25) are

$$\begin{aligned} m_{d,1}(y_t, y_{t-1}, \rho, \theta) &= \log q_t - a_1 - a_2 \log p_t && (29) \\ m_{d,2}(y_t, y_{t-1}, \rho, \theta) &= x_t m_{d,1}(y_t, y_{t-1}, \rho, \theta) && (30) \\ m_{d,3}(y_t, y_{t-1}, \rho, \theta) &= x_{t-1} m_{d,1}(y_t, y_{t-1}, \rho, \theta) && (31) \\ \theta &= (a_1, a_2) \\ \rho & \quad \text{not used} \end{aligned}$$

The prior for ρ is independent normal with location the unconstrained maximum likelihood estimates of (27) and scale twice the maximum likelihood standard deviation. The prior for $\theta = (a_1, a_2)$ is independent normal with means $(12, -2)$ and standard deviations $(2, 2)$. The support conditions on R and P of (28) are that diagonals of R must be positive, the first diagonal element P must be positive, and the eigenvalues of the companion matrix of Σ_{t-1} must be less than one in absolute value. In addition, a_1 must be positive and a_2 negative.

The Surface Sampling Algorithm, with moment conditions, prior, and support conditions as immediately above and with tuning parameters as in Gallant (2022) provides a chain of 50,000 draws after transients have died out. The chain was reduced (downsampled) with a stride of 10 leaving a chain of length 5,000 for analysis.

Figure 3 about here.

Table 2 about here.

Table 3 about here.

Table 4 about here.

Table 5 about here.

Table 6 about here.

Table 7 about here.

At the midpoint c of each edge in the graph $\mathcal{G}_\epsilon$, $\|q(c)\|$ was evaluated, where q is given by (2). Figure 3 plots the mean, 99th, and 90th quantiles of these $\|q(c)\|$ against the value of Δ that determined $\mathcal{G}_\epsilon$. The manifold M_ϵ appears to be moderately flat because the reduction from 50,000 to 5,000 draws, which increases the length of edges, leaves the centers of all edges reasonably close to M. As we shall see later in Subsection 4.3, a strongly curved manifold does not permit such a reduction.

The endpoints of Figure 3 are $\Delta = 0.9$, which is the smallest value for which M_ϵ is a connected set and $\Delta = 11$, which is the smallest value for which each point in $\{\mathsf{x}_i\}_{i=1}^N$ is connected by an edge to all other points in $\{\mathsf{x}_i\}_{i=1}^N$. Standard deviations for the two endpoints, $\Delta = 0.9$ and $\Delta = 11$, are presented in Tables 2 and 4, respectively, for all methods of computing standard deviations described in Section 3. Table 3 is the same for $\Delta = 3$, which is the point just after the curves in Figure 3 begin to flatten.

There are too many covariances among $\mathsf{x} = (\rho, \theta)$ to permit tabular display. But in an application, what is usually of interest are the covariances, expressed as a correlation, among the parameters θ introduced via the moment functions. These are displayed in Table 5.

Regressions among standard deviations are shown in Table 6 for $\Delta = 0.9$, $\Delta = 3.0$, and $\Delta = 11$. As seen from Table 6, for large Δ the estimators behave similarly, mostly making small adjustments to one another via the intercept term. For small Δ what stands out is the behavior of standard deviations from the modified Riemann estimator V_{MR}: standard deviations become larger relative to the other estimators. One can also see this behavior in Table 2.

Regressions among covariances are shown in Table 7. The story is much the same as for Table 6 except for scaling: Taking square roots to get standard deviations attenuates slopes.

If one regressed variances instead of standard deviations, the scaling would be similar. Standard deviations are of most interest in applications, hence the choice of standard deviations for Table 6 rather than variances.

For $\Delta = 3.0$ and standard errors from V_{ME}, the value of τ for which $P(R_\tau \,|\, x, y) = 0.95$ is $\tau = 2.64$; R_τ is defined by (20).

4.2 Extraction of the Stochastic Discount Factor

Consider the specification and extraction of the *ex post* stochastic discount factor.

Let $R_t = \frac{P_t + D_t}{P_{t-1}}$ denote the gross return to a security whose price is P_t at time t and that pays a dividend D_t at time t. Let $r_t = \log(P_t + D_t) - \log(P_{t-1})$ denote its geometric return. For any security, the stochastic discount factor satisfies

$$1 = \int \mathrm{SDF}_t(y)\, R_t(y)\, f(y \,|\, x_{t-1}, \rho)\, dy \tag{32}$$

provided that the SDF_t and R_t are functions of y. The density $f(y_t \,|\, x_{t-1}, \rho)$ is that given by (1) and x_{t-1} is the time $t-1$ information set of the conditional expectation (32).

Observed $y_{1,t}$ is daily, inflation adjusted, geometric returns on the S&P500 stock index (including distributions) and $y_{2,t}$ the same for the NASDAQ stock index for January 1, 2010, to December 31, 2018, which are $n = 2264$ bi-variate observations. The data are available at `www.aronaldg.org/webfiles/data` as `stocks_s.dat` for data and `stocks_s.doc` for documentation.

Consider a log quadratic specification of $sdf = \log(\mathrm{SDF})$; viz.,

$$sdf_q(y_t) \;\; = \;\; a_0 + a_1 f_t + a_2 f_t^2, \tag{33}$$

where $f_t = \frac{1}{2}(y_{1,t}/100 + y_{2,t}/100)$. The posterior probability in favor of a quadratic rather than linear specification is 0.81 (Gallant, 2022, Subsection 3.2).

For the likelihood (1) we use a bivariate SNP-ARCH model. This model has an SNP innovation density with VAR location and diagonal ARCH scale. It is parameterized as follows:

$$y \sim \frac{[\mathcal{P}(z)]^2 n_2(y \,|\, \mu, \Sigma)}{\int [\mathcal{P}(s)]^2 n_2(s \,|\, 0, I)\, ds} \tag{34}$$

where $\mathcal{P}(z)$ is evaluated at $z = \Sigma^{-1/2}(y - \mu)$, $n_2(y|\,\mu, \Sigma)$ is the bivariate normal density, and

$$\mathcal{P}(z) = a_{01}z_2 + a_{02}z_2^2 + a_{03}z_2^3 + a_{04}z_2^4 + a_{05}z_1 + a_{06}z_1^2 + a_{07}z_1^3 + a_{08}z_1^4.$$

The location and scale are

$$\mu = b_0 + By_{t-1}$$

$$\Sigma = R_0R_0' + P(y_{t-1} - b_0 - By_{t-2})(y_{t-1} - b_0 - By_{t-2})'P'$$

where R_0 is upper triangular and P is diagonal. The parameters of $\mathcal{P}(z)$, the elements of b_0 and B, and the non-zero elements of R_0 and P are the elements of ρ; see Table 8 for the ordering of parameters within ρ. The information set is $x_{t-1} = (y_{t-1}, y_{t-2})$.

The moment conditions defining (2) that we use in estimation are:

$$\begin{aligned}
m_1(y_t, x_{t-1}, \rho, \theta) &= 1.0 - \exp[sdf_q(y_t) + r_{1,t}] && (35)\\
m_2(y_t, x_{t-1}, \rho, \theta) &= 1.0 - \exp[sdf_q(y_t) + r_{2,t}] && (36)\\
m_3(y_t, x_{t-1}, \rho, \theta) &= y_{1,t-1}m_1(y_t, x_{t-1}, \rho, \theta) && (37)\\
m_4(y_t, x_{t-1}, \rho, \theta) &= y_{2,t-1}m_2(y_t, x_{t-1}, \rho, \theta) && (38)\\
r_t &= y_t/100\\
\theta &= (a_0, a_1, a_2)\\
\rho &\quad \text{not used}
\end{aligned}$$

The prior for ρ is independent normal with location and scale the SNP-ARCH unconstrained maximum likelihood estimated parameters and standard errors. Admittedly this is a data dependent, independence prior, but it is so loose that we think this consideration can be dismissed. The prior for $\theta = (a_0, a_1, a_2)$ is independent normal with means $(0, -1, 0)$ and standard deviations $(1, 1, 1)$. This prior loosely implies a variant of CAPM (capital asset pricing model).

The support conditions apply to R_0 and P of the scale function and B of the location function. They are that the diagonals of R_0 be positive and that the first diagonal element of the diagonal matrix P be positive. In addition, the eigenvalues of the companion matrices for the location function and the scale function are required to be less than one in absolute value.

The Surface Sampling Algorithm, with moment conditions, prior, and support conditions as immediately above and with tuning parameters as in Gallant (2022) provides a chain of 50,000 draws after transients have died out. The chain was reduced with a stride of 10 leaving a chain of length 5,000 for analysis.

Figure 4 about here.

Table 8 about here.

Table 9 about here.

Table 10 about here.

Table 11 about here.

Table 12 about here.

The manifold M_ϵ appears to be almost Euclidean as seen in Figure 4: The midpoints of all edges are in M_ϵ to within a tolerance of 2×10^5 for any Δ. The construction of Figure 4 the same as described above for Figure 3. Here, $\Delta = 1$ is the smallest value for which M_ϵ is a connected set and $\Delta = 31$ the smallest value for which each node is connected to every other node by an edge. Results for $\Delta = 1$ are rather erratic; thus $\Delta = 2$ is the smallest value for which results are presented.

Standard deviations for $\Delta = 2$ and $\Delta = 31$ are presented in Tables 8 and 10 respectively for all methods of computing standard deviations described in Section 3. Table 9 is the same for $\Delta = 10$, which is the point just after the curves in Figure 4 begin to flatten. The covariances expressed as correlation for the parameters θ that enter the model exclusively through the moment conditions are shown in Table 5.

Regressions among standard deviations are shown in Table 11 for $\Delta = 2$, $\Delta = 10$, and $\Delta = 31$. As seen from Table 11, for large Δ the estimators behave similarly, mostly

making small adjustments to one another via the intercept term. For small Δ results differ dramatically. Consider the behavior of V_{MR} relative to either V_{EC} or V_{IC}, results being essentially the same for either of the latter. Standard errors computed from V_{MR} are approximately doubled.

Regressions among covariances are shown in Table 12. The story is much the same as for Table 11 except for scaling for the reasons discussed above.

For $\Delta = 10.0$ and standard errors from V_{ME}, the value of τ for which $P(R_\tau \,|\, x, y) = 0.95$ is $\tau = 2.85$; R_τ is defined by (20).

The examples in Subsections 4.1 and 4.2, while representative of actual applications, are deficient in that M is too flat, i.e., too like an Euclidean space. We need an example with more curvature to gain further insight as to how the four measures of scale behave. This we consider in the next example.

4.3 A Curved Manifold Example

The curved manifold example has a bivariate normal, iid likelihood

$$y_t \sim n_2(y_t \,|\, \mu, \Sigma) \tag{39}$$

$$\Sigma = RR', \tag{40}$$

where R is upper triangular. Thus,

$$\rho = (\mu_1, \mu_2, R_{1,1}, R_{1,2}, R_{2,2}) \in \mathbb{R}^5.$$

The moment conditions are

$$m_{c,1}(y_t, y_{t-1}, \rho, \theta) = y_{1,t}^2 + y_{2,t}^2 - 4\theta \tag{41}$$

$$m_{c,2}(y_t, y_{t-1}, \rho, \theta) = (y_{1,t} - y_{1,t-1})^2 - 2\theta \tag{42}$$

$$\theta \in \mathbb{R}^1$$

$$\rho \qquad \text{not used}$$

Note that the moment functions depend on a lag even though the data are iid. The data, $n = 500$, are simulated with $\mu_1 = 0$, $\mu_2 = 0$, $\Sigma_{1,1} = 5$, $\Sigma_{1,2} = \Sigma_{2,1} = 6.12372$, $\Sigma_{2,2} = 15$, and $\theta = 5$. The prior for ρ is independent normal with location the unconstrained maximum

likelihood estimates of (39) and standard deviation 5.0. The prior for θ is normal with mean 5.0 and standard deviation 5.0. The support conditions are that diagonals of R must be positive and θ must be positive.

Taking expectations, the conditions on ρ and θ are

$$\Sigma_{1,1} + \Sigma_{2,2} + \mu_1^2 + \mu_2^2 = 4\theta \tag{43}$$

$$\Sigma_{1,1} = \theta \tag{44}$$

$$\tag{45}$$

Figure 5 displays this surface.

Figure 5 about here.

Figure 6 about here.

The Surface Sampling Algorithm, with moment conditions, prior, and support conditions as immediately above provided a chain of 50,000 draws after transients have died out. The chain was reduced with a stride of 10 leaving a chain of length 5,000.

For the reduced (downsampled) chain of length 5000, the smallest value of Δ such that M_ϵ is a connected set is $\Delta = 0.63$. The norms of q given by (2) evaluated at the centers of the edges of the graph $\mathcal{G}_\epsilon$ had mean, 99th percentile, etc. on the order of 1×10^{-2}; the edges of $\mathcal{G}_\epsilon$ are not comfortably within M.

For the full chain of length 50,000, the smallest value of Δ such that M_ϵ is a connected set is $\Delta = 0.57$. $\Delta = 15$ the smallest value for which each point in $\{\mathsf{x}_i\}_{i=1}^N$ is connected by an edge to all other points in $\{\mathsf{x}_i\}_{i=1}^N$. The norms of q evaluated at the centers of the edges of the graph $\mathcal{G}_\epsilon$ had mean, 99th percentile, etc. of the order of 1×10^{-3}, see Figure 6: the edges are marginally within M. In the complete chain of length 50,000 the separation between successive points is determined by the proposal density of the Surface Sampling Algorithm, which, for these draws, is $N_6(0, 2^{-6}I)$. Thus, the graph $\mathcal{G}_\epsilon$ at $\Delta = 0.57$ consists mostly of edges between successive points of the complete chain (that has had repetitions due to rejections removed). One can reduce the average distance of edge midpoints from q by decreasing the variance of the proposal density.

In this instance the dimension of $\mathsf{x} = (\rho, \theta)$ is small enough that all scale estimates can be displayed. Standard deviations and correlations for the two endpoints of Figure 6, $\Delta = 0.57$ and $\Delta = 15$, are presented in Tables 13 and 15, respectively, for all methods of computing standard deviations described in Section 3. Table 14 is the same for $\Delta = 3.0$, which is the point just after the curves in Figure 4 begin to flatten.

Table 13 about here.

Table 14 about here.

Table 15 about here.

Table 16 about here.

Table 17 about here.

Table 16 presents the results of simple regressions among the standard deviations from V_{EC}, V_{IC}, V_{ME}, and V_{MR} for $\Delta = 0.57$, $\Delta = 3$, and $\Delta = 15$. As was the case for the previous examples, for large Δ the estimators behave similarly. The intercept terms for each are negligible relative to standard deviations from the others and slopes are near one. For small Δ it is standard deviations from V_{ME} that behave differently from the others rather than standard deviations from V_{MR} as was the case in Table 11.

Regressions among covariances are shown in Table 17. The story is much the same as for Table 16 except for scaling for the reasons discussed above.

For $\Delta = 10.0$ and standard errors from V_{ME}, the value of τ for which $P(R_\tau \,|\, x, y) = 0.95$ is $\tau = 1.69$; R_τ is defined by (20).

5 Interpretation

The reference system for the MCMC chain is extrinsic; that is the parameters $\mathsf{x} = (\rho, \theta)$ are expressed in a d_a dimensional Euclidean coordinate system rather than a d dimensional coordinate system defined either on the manifold M or on a chart of M. Therefore, if there is an interest in credibility intervals, rectangles, or regions, then, presumably, these would be expressed relative to the extrinsic coordinate system.

Geometrically, what one does to construct a credibility interval, rectangle, or region is to shift the origin of the coordinate system to either the extrinsic mean $\tilde{\mathsf{x}}$ or the intrinsic mean $\bar{\mathsf{x}}$ then consider a geometrical shape centered at this mean. Therefore it is possible that a region constructed using the extrinsic mean, in particular the rectangle

$$\tilde{R}_\tau = \times_{i=1}^{d_a} [\tilde{\mathsf{x}}_i - \tau \, \mathrm{sdev}(\mathsf{x}_i), \; \tilde{\mathsf{x}}_i + \tau \, \mathrm{sdev}(\mathsf{x}_i)],$$

will not contain any points on the manifold. This would seem to be a compelling argument in favor of using V_{IC}, V_{ME}, or V_{MR}. We will adopt that view here thus ruling out the use of V_{EC}.

Computation of the intrinsic mean requires a choice of Δ. Fortunately, the intrinsic mean is apparently not sensitive to the choice of Δ. In our examples, the intrinsic mean does not change at all for Δ to the right of the point where the midpoint curve flattens, cf. Figure 3, 4, or 6. For the curved manifold example of Subsection 4.3, it is the same regardless of Δ. For the others, changes in the intrinsic mean are not dramatic for Δ to the left of the point where the midpoint curve flattens

The main attraction of V_{IC} is that it is straightforward to compute and does not require application of Dijkstra's algorithm[10] nor computation of the tangent space coordinate system $T_{\bar{\mathsf{x}}}$. The regressions in Tables 6, 11, and 11 imply that it is reasonable relative to the other estimators for Δ to the right of the point where the midpoint curve flattens. For Δ to the left of the point where the midpoint curve flattens a choice involves the following considerations.

The logic of the estimator V_{ME} seems compelling with respect to standard deviations. A standard deviation pertains to one axis of a Euclidean coordinate system. What matters is how far to the left or right of the chosen measure of location an MCMC draw x_i lies along that

[10] Other than to compute the intrinsic mean.

axis. Instead of V_{ME} increasing that distance according to the entire length of the geodesic $\delta(\mathsf{x}_i, \bar{\mathsf{x}})$, as does V_{MR}, it only increases it by how far the geodesic moves along the relevant axis. This seems compelling because it makes little sense to increase length along an axis by the entire length of a geodesic when the movement in the geodesic is mostly orthogonal to that axis. Unfortunately, the interpretation of a correlation or covariance computed from V_{ME} is less straightforward due to the unequal scaling of coordinates.

The Riemann approach is somewhat standard in the present context (Pennec, 2006). However, its target application is the characterization of the probability measure P in the probability space $(M, \mathcal{B}, P)$ when one has an analytic, differentiable expression for geodesics, can easily find the geodesics passing through any given point $\mathsf{x} \in M$, and for which there is a convenient one-to-one mapping between the chart $T_{\bar{\mathsf{x}}}M$ and M. Under these conditions, one can either define or deduce from data a distribution on the Euclidean space $T_{\bar{\mathsf{x}}}M$ using standard statistical methods and then easily transfer it to M. We are concerned here with the case that the distribution on the manifold M is known, at least as far as to be able to generate MCMC chains for x on the manifold, and what is required is some reasonable notion of the scale of the fluctuations of x in the chain. Our adaption of the Riemann geometry concepts to this context is reasonable. In fact, the chart developed here appears to have more structural appeal for this purpose than the conventional Riemann chart, were it applicable. The modified Riemann estimator has one distinctive feature not shared by the others: V_{MR} is singular with rank d as is the manifold M. Except in rare instances, such as the manifold M is a flat space and the MCMC algorithm places points on M with high precision, the other estimators will not share this feature. However, due to the singularity and the nonlinearity of M one is quite likely underestimating the variation along some of the extrinsic axes, cf. the last panel of Table 1.

The estimators V_{ME} and V_{MR} can behave erratically when computed for Δ to the left of the point where the midpoint curve (e.g., Figure 3) flattens. The argument that justifies replacing M by M_ϵ (Memoli and Sapiro, 2001) for the purpose of computing geodesics is an asymptotic argument as Δ tends to zero. So it would seem that the largest Δ one would prefer is slightly to the right of the point where the midpoint curve flattens when using V_{ME} or V_{MR}.

The view here is that V_{ME} makes the most intuitive sense for tabular reporting of estimation results and for constructing credibility intervals or rectangles such as (20) and that V_{IC} is better for general use because its covariances and correlations are easier to interpret.

6 Conclusion

This paper studies approaches to estimating a variance matrix from an MCMC chain on a curved, nonlinear manifold and their use together with the chain for the construction of credibility intervals, rectangles, and regions. Four notions are examined: extrinsic centered at the extrinsic mean, denoted V_{EC}, extrinsic centered at the intrinsic mean, denoted V_{IC}, the estimator V_{ME}, which is V_{IC} modified by adjusting distance from the intrinsic mean upward to reflect geodesic distance, and the estimator V_{MR}, which is Riemannian variance and covariance modified to reflect the practicalities of having only an MCMC chain and tangent plane rather than a complete analytic description of the manifold. Conceptually one would prefer one of the latter two, V_{ME} or V_{MR}. They are computed by placing ϵ-balls around the MCMC chain to get a connected set M_ϵ. Because posterior regions are not balls, used instead of ϵ-balls are rectangles which are determined from the MCMC chain that are indexed by Δ instead of ϵ. Geodesics are computed from a graph connecting all pairs within M_ϵ that are no farther apart than Δ. The value of Δ is determined by examining a plot of quantiles of the distance to the manifold of the centers of lines between connected pairs in the graph, cf. Figure 3, 4, or 6. The advice of this paper is to choose Δ just to the right of the point where the curve flattens, use V_{ME} for reporting standard deviations of estimation results and for computing credibility intervals and rectangles such as (20). Further advice is to use V_{IC} for credibility regions that make use of covariance or any other computation that uses covariance or correlation.

7 References

Bornn, Luke, Neil Shephard, and Reza Solgi (2018), "Moment Conditions and Bayesian Nonparametrics," *Journal of the Royal Statistical Society, Series B* 81(1), 5–43.

Businger, Peter A., and Gene H. Golub (1969), "Singular Value Decomposition of a Complex

Matrix," *Communications of the ACM* 12, 564–565.

Byrne, S., and M. Girolami (2013), "Geodesic Monte Carlo on Embedded Manifolds," *Scandinavian Journal of Statistics* 40, 825–845.

Dijkstra, E. W. (1959), "A Note on Two Problems in Connexion with Graphs," *Numerische Mathematik* 1 269–271

Gallant, A. Ronald (2022), "Nonparametric Bayes Subject to Overidentified Moment Conditions," *Journal of Econometrics* 228, 27–38.

Gallant, A. Ronald, and Douglas W. Nychka (1987), "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica* 55, 363–390.

Gallant, A. Ronald, and George Tauchen (1989), "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica* 57, 1091–1120.

Memoli, Facundo, and Guillermo Sapiro (2001), "Fast Computation of Weighted Distance Functions and Geodesics on Implicit Hyper-surfaces," *Journal of Computational Physics* 173, 730–764.

Morgan, Frank (2016), *Geometric Measure Theory, A Beginner's Guide, 5th Edition,* San Diego CA, Academic Press.

Pennec, Xavier (1999), "Probabilities and Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. In A.E. Cetin, L. Akarun, A. Ertuzun, M.N. Gurcan, and Y. Yardimci, editors, *Proceedings of Nonlinear Signal and Image Processing* 1, 194–198, June 20-23, Antalya, Turkey, 1999. IEEE-EURASIP.

Pennec, Xavier (2006), "Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements," *Journal of Mathematical Imaging and Vision* 25, 127-154.

Sethian, J. A. (1996), "A Fast Marching Level Set Method for Monotonically Advancing Fronts," *Proceedings of the National Academy of Sciences* 93, 1591–1595. https://doi.org/10.1073/pnas.93.4.1591

Shin, Minchuyl (2015), "Bayesian GMM," Working paper, Department of Economics, University of Illinois. http://www.econ.uiuc.edu/∼mincshin/BGMM_ver05

Wikipedia (2021), "Frechet Mean," Last modified June 3, 2021. https://en.wikipedia.org/wiki/Frechet_mean.

Zappa, Emilio, Miranda Holmes-Cerfon,and Jonathan Goodman (2018), "Monte Carlo on Manifolds: Sampling Densities and Integrating Functions," *Communications on Pure and Applied Mathematics* 71, 2609–2647.

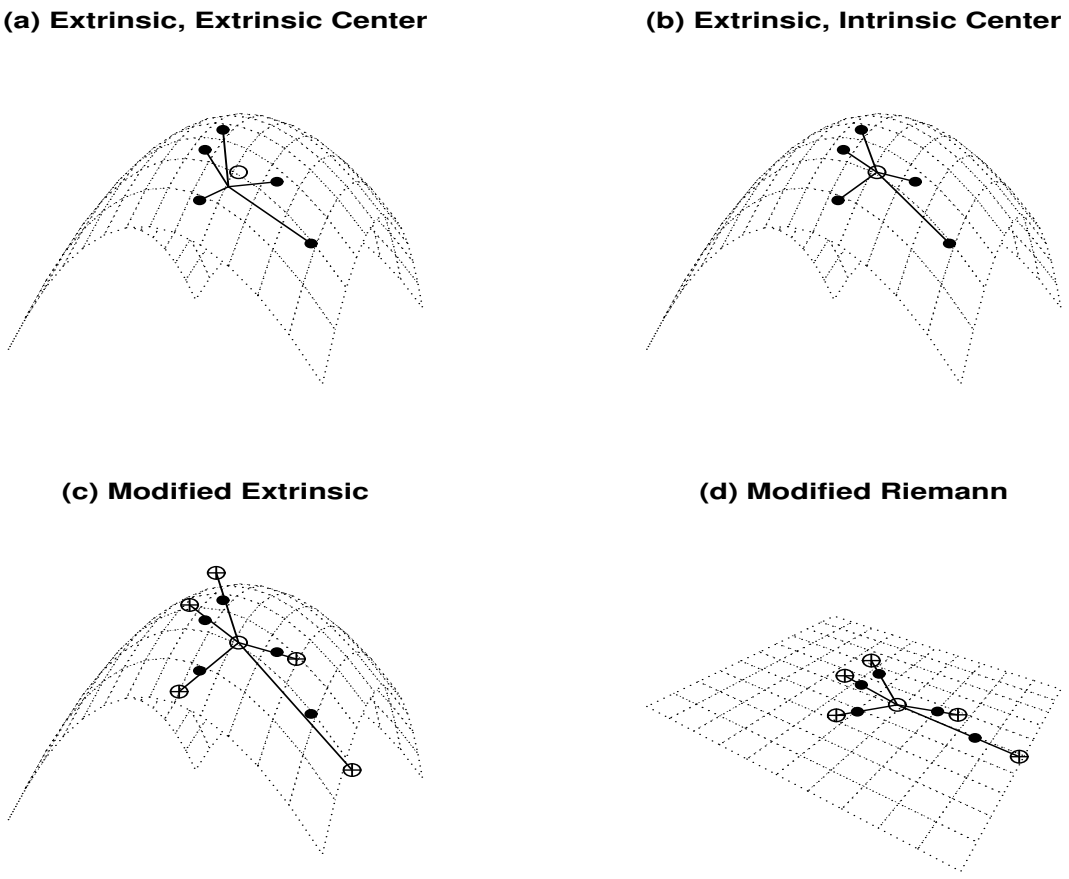

Figure 1. Illustration of Scale Measures Panels (a), (b), and (c) show a hypothetical surface with hypothetical sample points shown as solid dots, ●, and the intrinsic mean, $\bar{\mathrm{x}}$, shown as an open circle, ○. Panel (d) shows the plane tangent to the hypothetical surface at the intrinsic mean with the solid dots and open circle projected onto that plane.

In Panel (a) are vectors formed by connecting the extrinsic mean, $\tilde{\mathrm{x}}$, to the sample points, ●. The scale measure V_{EC} is the average of the outer product of these vectors. This is the standard measure of scale, S^2, for any sample.

In Panel (b) are vectors formed by connecting the intrinsic mean, $\bar{\mathrm{x}}$, to the sample points, The scale measure V_{IC} is the average of the outer product of these vectors.

In Panel (c) are vectors formed by extending the vectors of Panel (b) by the length of their geodesics, coordinate by coordinate, to connect to the points shown as circled pluses, $\oplus$. Because the multiples of coordinates can differ, the circled plus vectors need not pass through the sample points. The scale measure V_{ME} is the average of the outer product of the circled plus vectors.

In Panel (d) are vectors on the tangent plane $T_{\bar{\mathrm{x}}}M$ that are formed by extending the vectors connecting the intrinsic mean, ○. to the projected sample points, ● to the points shown as circled pluses, $\oplus$, by the length of the geodesics connecting ○ to ● on the manifold M. The circled plus vectors will pass through the projected sample points. The scale measure V_{MR} is the average of the outer product of the circled plus vectors.

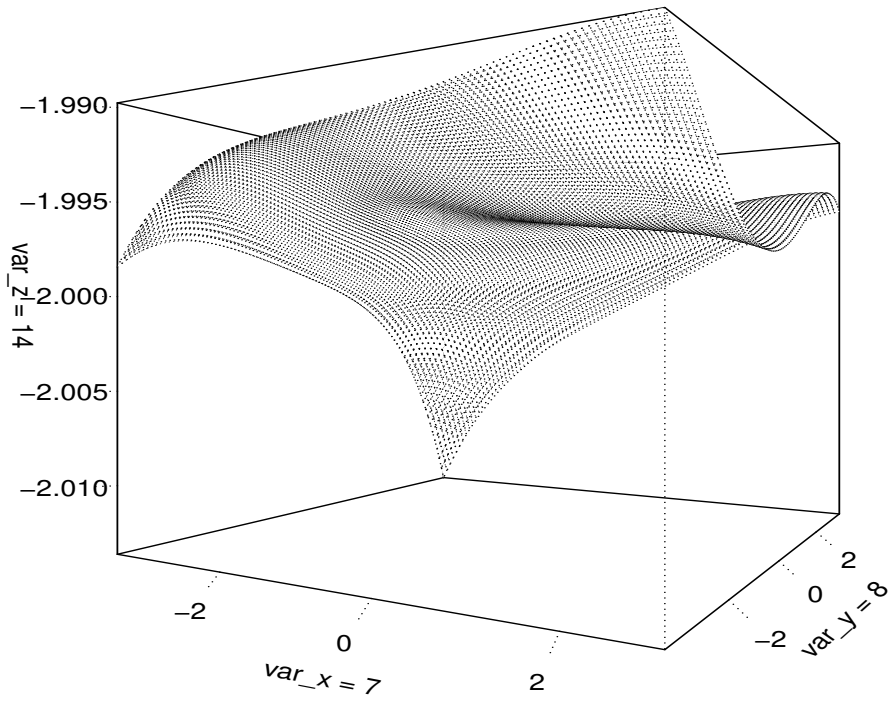

Figure 2. 95% Credibility Region, Demand and Supply Example. The demand and supply example is described in Subsection 4.1. The coordinates `var_x` and `var_y` are the seventh and eighth chart coordinates, that is, they are the coefficients of the seventh and eighth columns of $T_{\bar{\mathsf{x}}}^{\vee}$; `var_z` is the last element of $\mathsf{x} = (\rho, \theta)$. It is the price elasticity of demand. All other chart coordinates are held fixed at the values of the intrinsic mean. The surface was obtained by fitting a multivariate polynomial of degree four with draws $\{\mathsf{x}_i\}$ as the dependent variable and the corresponding points $\{z_i\}$ on the chart as the independent variable. In this instance, `var_x` is roughly interpretable as ρ_{10}, which is $P_{1,1}$, and `var_y` is roughly interpretable as ρ_{11}, which is $P_{2,2}$. $P_{1,1}$ and $P_{2,2}$ are the parameters that determine the stochastic volatility of log price and log quantity, respectively.

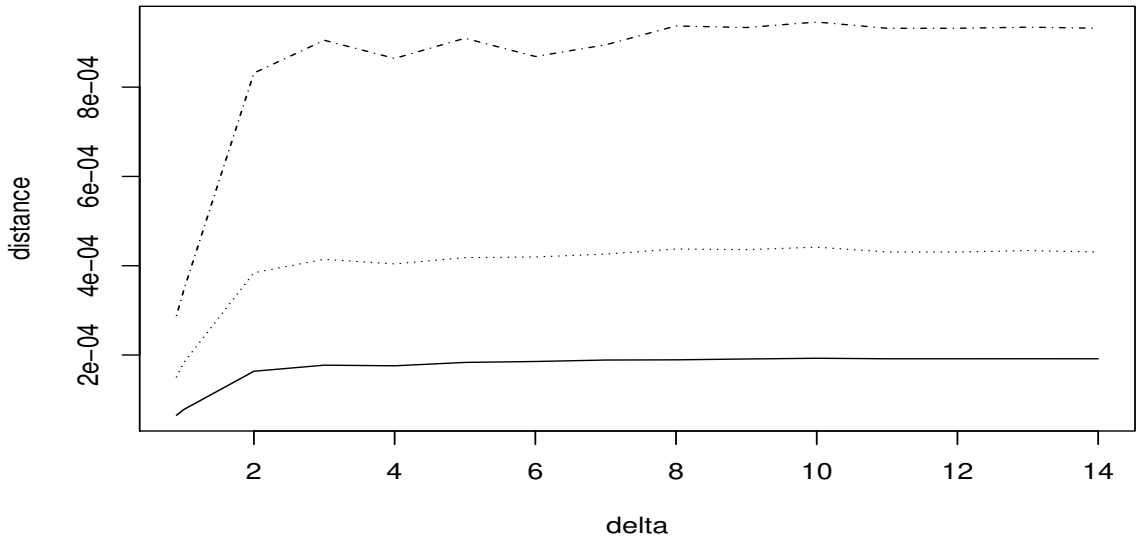

Figure 3. Distance of Edge Midpoints from Manifold, Demand and Supply Example. For the graph $\mathcal{G}_\epsilon$ with offset Δ as shown on the horizontal axis, the distance of the center of each edge from the manifold M is computed. The dotdash line is the 99th percentile, the dotted line is the 90th percentile, and the solid line is the mean.

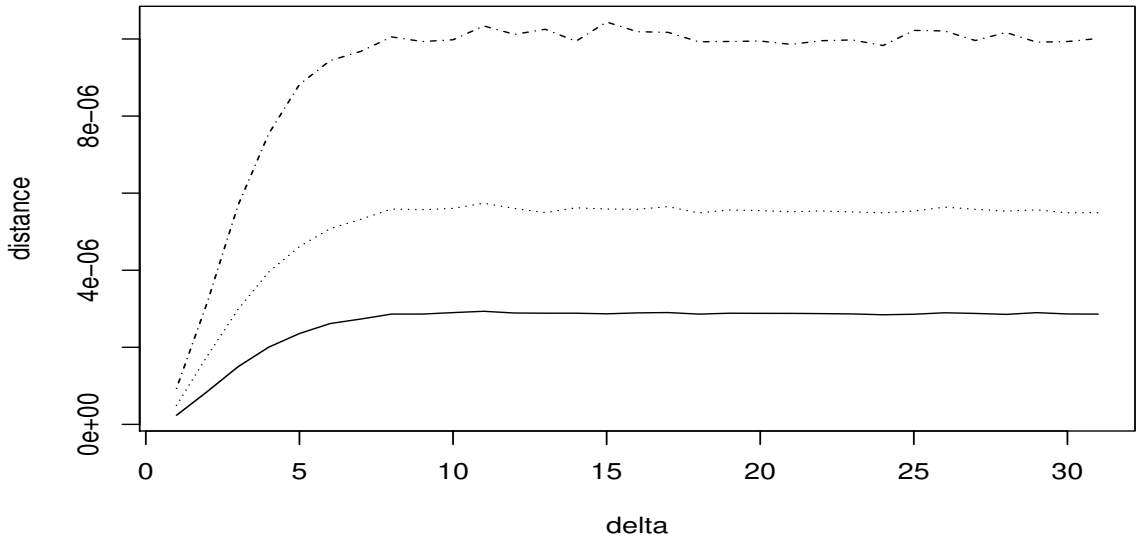

Figure 4. Distance of Edge Midpoints from Manifold, Stochastic Discount Function Example For the graph $\mathcal{G}_\epsilon$ with offset Δ as shown on the horizontal axis, the distance of the center of each edge from the manifold M is computed. The dotdash line is the 99th percentile, the dotted line is the 90th percentile, and the solid line is the mean.

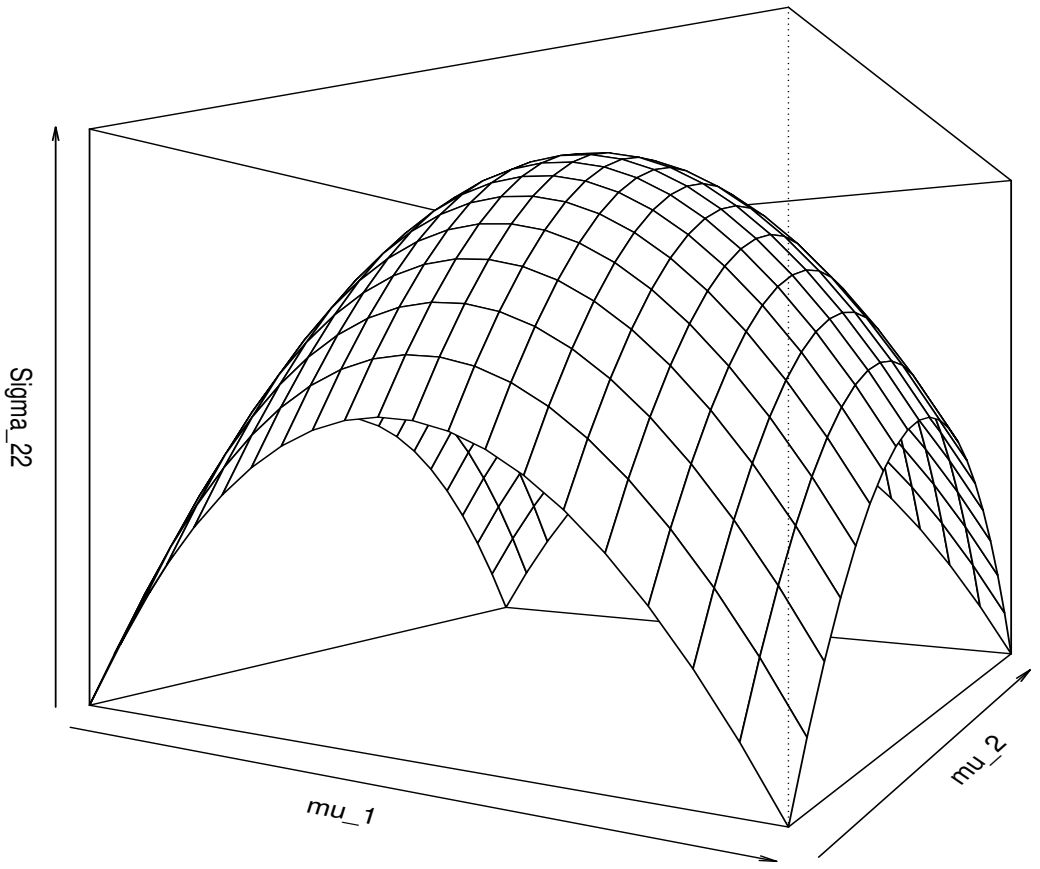

Figure 5. Curved Manifold Example Plotted is the manifold M for the likelihood (39) subject to moment conditions (2) determined by (41) and (42). The missing dimensions, $\Sigma_{1,1}$, $\Sigma_{1,2}$, and θ, are held constant at 5, 6.12372, and 5, respectively.

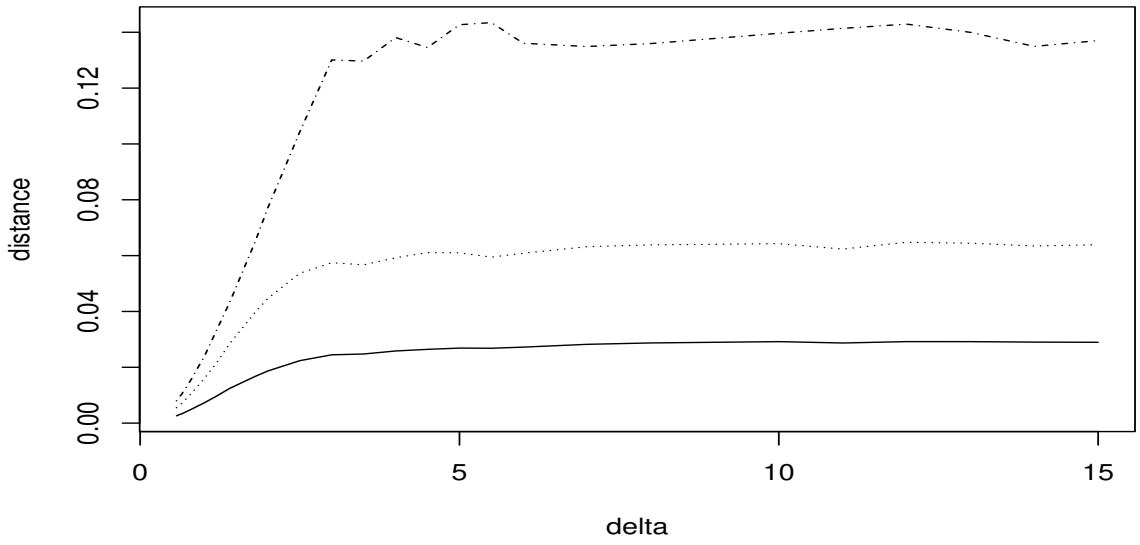

Figure 6. Distance of Edge Midpoints from Manifold, Curved Manifold Example For the graph $\mathcal{G}_\epsilon$ with offset Δ as shown on the horizontal axis, the distance of the center of each edge from the manifold M is computed. The dotdash line is the 99th percentile, the dotted line is the 90th percentile, and the solid line is the mean.

Table 1. Illustration of Population Variances and Correlations

V_{EC}			C_{EC}		
0.073658	0.058105	0.000088	1.000000	0.788701	0.002893
0.058105	0.073686	0.000041	0.788701	1.000000	0.001337
0.000088	0.000041	0.012574	0.002893	0.001337	1.000000
V_{IC}			C_{IC}		
0.073658	0.058105	0.000083	1.000000	0.788701	0.002194
0.058105	0.073686	0.000034	0.788701	1.000000	0.000896
0.000083	0.000034	0.019536	0.002194	0.000896	1.000000
V_{ME}			C_{ME}		
0.073658	0.058105	0.000093	1.000000	0.788701	0.002262
0.058105	0.073686	0.000037	0.788701	1.000000	0.000913
0.000093	0.000037	0.022778	0.002262	0.000913	1.000000
V_{MR}			C_{MR}		
0.073658	0.058105	0.000000	1.000000	0.788701	0.000000
0.058105	0.073686	0.000000	0.788701	1.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Shown are the variance matrices V_{EC}, V_{IC}, V_{ME}, and V_{MR} and correlation matrices C_{EC}, C_{IC}, C_{ME}, and C_{MR} computed from a simulation of (16) of length $n = 198373$.

Table 2. Demand and Supply Example, $\Delta = 0.9$

	Mean		Standard Deviation			
			Extrinsic		Modified	
Parameter	Extrinsic	Intrinsic	Extr Ctr	Intr Ctr	Extrinsic	Riemann
μ_1	0.006974	-0.000014	0.032410	0.033155	0.056276	0.000416
μ_2	-0.006384	-0.007252	0.034833	0.034843	0.078493	0.055205
μ_3	-0.001982	0.007796	0.035280	0.036610	0.081187	0.057796
$R_{1,1}$	0.995638	0.985301	0.030594	0.032293	0.072945	0.051602
$R_{1,2}$	-0.000188	-0.009587	0.019377	0.021537	0.061537	0.033760
$R_{2,2}$	1.001946	1.050761	0.031913	0.058325	0.093115	0.090660
$R_{1,3}$	-0.004291	-0.003850	0.018963	0.018969	0.054806	0.030062
$R_{2,3}$	0.001238	-0.006318	0.018397	0.019889	0.050946	0.031336
$R_{3,3}$	0.996106	0.984606	0.030197	0.032314	0.076838	0.051241
$P_{1,1}$	0.137964	0.173574	0.081201	0.088667	0.113114	0.142283
$P_{2,2}$	0.004711	-0.012474	0.109165	0.110509	0.123758	0.177366
$P_{3,3}$	-0.058212	-0.058302	0.128347	0.128347	0.138672	0.203544
a_1	11.986857	11.982728	0.010649	0.011422	0.028206	0.018471
a_2	-1.996886	-1.994403	0.006776	0.007217	0.018729	0.011906

The data are a simulation of the demand and supply system (24) through (26). An MCMC chain of length 50,000 was computed using the Surface Sampling Algorithm for the likelihood (27) subject to moment conditions (2) as determined by (29) through (31) The prior for ρ is independent normal with location the unconstrained maximum likelihood estimates of (27) and scale twice the maximum likelihood standard errors. The prior for $\theta = (a_1, a_2)$ is independent normal with means $(12, -2)$ and standard deviations $(2, 2)$. The support conditions on R and P of (28) are that diagonals of R must be positive, the first diagonal element P must be positive, and the eigenvalues of the companion matrix of Σ must be less than one in absolute value. In addition, a_1 must be positive and a_2 negative. The chain was reduced (downsampled) with a stride of 10 leaving a chain of length 5,000 for computations. Means and standard deviations shown in the table for offset $\Delta = 0.9$, which is the smallest value of Δ for which the manifold M_ϵ is connected.

Table 3. Demand and Supply Example, $\Delta = 3$

	Mean		Standard Deviation			
			Extrinsic		Modified	
Parameter	Extrinsic	Intrinsic	Extr Ctr	Intr Ctr	Extrinsic	Riemann
μ_1	0.006974	-0.000005	0.032410	0.033154	0.034008	0.000024
μ_2	-0.006384	0.018046	0.034833	0.042547	0.044250	0.043604
μ_3	-0.001982	0.013853	0.035280	0.038671	0.040339	0.039671
$R_{1,1}$	0.995638	0.999636	0.030594	0.030854	0.032664	0.031616
$R_{1,2}$	-0.000188	-0.001498	0.019377	0.019421	0.022899	0.019980
$R_{2,2}$	1.001946	1.021762	0.031913	0.037566	0.038724	0.038680
$R_{1,3}$	-0.004291	-0.030309	0.018963	0.032197	0.033586	0.033113
$R_{2,3}$	0.001238	0.021554	0.018397	0.027409	0.029031	0.027764
$R_{3,3}$	0.996106	0.967894	0.030197	0.041327	0.043449	0.042494
$P_{1,1}$	0.137964	0.168808	0.081201	0.086863	0.087303	0.089219
$P_{2,2}$	0.004711	-0.001004	0.109165	0.109314	0.109444	0.114385
$P_{3,3}$	-0.058212	-0.052750	0.128347	0.128463	0.128560	0.133870
a_1	11.986857	11.991603	0.010649	0.011659	0.012774	0.012461
a_2	-1.996886	-1.998937	0.006776	0.007080	0.007853	0.007601

As for Table 2 except that $\Delta = 3$, which is the point just after the curves in Figure 3 begin to flatten.

Table 4. Demand and Supply Example, $\Delta = 11$

	Mean		Standard Deviation			
			Extrinsic		Modified	
Parameter	Extrinsic	Intrinsic	Extr Ctr	Intr Ctr	Extrinsic	Riemann
μ_1	0.006974	-0.000005	0.032410	0.033154	0.033150	0.000024
μ_2	-0.006384	0.018046	0.034833	0.042547	0.042543	0.043143
μ_3	-0.001982	0.013853	0.035280	0.038671	0.038668	0.039171
$R_{1,1}$	0.995638	0.999636	0.030594	0.030854	0.030851	0.031218
$R_{1,2}$	-0.000188	-0.001498	0.019377	0.019421	0.019419	0.019721
$R_{2,2}$	1.001946	1.021762	0.031913	0.037566	0.037563	0.038075
$R_{1,3}$	-0.004291	-0.030309	0.018963	0.032197	0.032194	0.032654
$R_{2,3}$	0.001238	0.021554	0.018397	0.027409	0.027407	0.027407
$R_{3,3}$	0.996106	0.967894	0.030197	0.041327	0.041323	0.041930
$P_{1,1}$	0.137964	0.168808	0.081201	0.086863	0.086854	0.088216
$P_{2,2}$	0.004711	-0.001004	0.109165	0.109314	0.109303	0.110615
$P_{3,3}$	-0.058212	-0.052750	0.128347	0.128463	0.128451	0.129637
a_1	11.986857	11.991603	0.010649	0.011659	0.011658	0.012298
a_2	-1.996886	-1.998937	0.006776	0.007080	0.007079	0.007500

As for Table 2 except that $\Delta = 11$, which is the smallest value such that each node in M_ϵ is connected to all other nodes.

Table 5. Moment Function Parameter Correlations

	Extrinsic		Modified	
Correlation	Extr Ctr	Intr Ctr	Extrinsic	Riemann
	Demand and Supply Example, $\Delta = 0.9$			
$\rho(a_1, a_2)$	-0.953464	-0.959086	-0.850522	-0.970201
	Demand and Supply Example, $\Delta = 11$			
$\rho(a_1, a_2)$	-0.953464	-0.951455	-0.951455	-0.959869
	Stochastic Discount Factor Example, $\Delta = 2$			
$\rho(a_1, a_2)$	-0.538199	-0.624957	-0.619447	-0.515152
$\rho(a_1, a_3)$	-0.933849	-0.924537	-0.889809	-0.238756
$\rho(a_2, a_3)$	0.238997	0.341350	0.470240	0.442594
	Stochastic Discount Factor Example, $\Delta = 31$			
$\rho(a_1, a_2)$	-0.538199	-0.522063	-0.522063	-0.384290
$\rho(a_1, a_3)$	-0.933849	-0.931478	-0.931478	-0.991255
$\rho(a_2, a_3)$	0.238997	0.214916	0.214916	0.262880

Shown are the correlations for the parameters θ that appear in the moment functions (2) computed from V_{EC}, V_{IC}, V_{ME}, V_{MR} that were themselves computed from the MCMC chains described in Tables 2, 4, 8 and 10 for the four blocks of the table, respectively, as indicated by the headings for each block. For instance, the first entry $\rho(a_1, a_2) = -0.953464$ refers to a correlation computed from V_{EC} for the demand and supply MCMC chain described in Table 2.

Table 6. Regressions Among Standard Deviations, Demand and Supply Example

Variable				
Independent	Dependent	Intercept	Slope	R^2
		$\Delta = 0.9$		
V_{EC} sdev	V_{IC} sdev	0.003445	0.996188	0.966189
V_{EC} sdev	V_{ME} sdev	0.039708	0.837890	0.829003
V_{EC} sdev	V_{MR} sdev	0.000226	1.619744	0.914482
V_{IC} sdev	V_{ME} sdev	0.036227	0.853974	0.884487
V_{IC} sdev	V_{MR} sdev	-0.005547	1.629730	0.950897
V_{ME} sdev	V_{MR} sdev	-0.060075	1.713394	0.866591
		$\Delta = 3.0$		
V_{EC} sdev	V_{IC} sdev	0.005572	0.966685	0.985068
V_{EC} sdev	V_{ME} sdev	0.007508	0.951912	0.984039
V_{EC} sdev	V_{MR} sdev	0.002167	1.027389	0.922289
V_{IC} sdev	V_{ME} sdev	0.002007	0.985043	0.999616
V_{IC} sdev	V_{MR} sdev	-0.004049	1.069157	0.947510
V_{ME} sdev	V_{MR} sdev	-0.006268	1.086251	0.949377
		$\Delta = 11.0$		
V_{EC} sdev	V_{IC} sdev	0.005572	0.966685	0.985068
V_{EC} sdev	V_{ME} sdev	0.005572	0.966685	0.985068
V_{EC} sdev	V_{MR} sdev	0.002657	0.993814	0.919170
V_{IC} sdev	V_{ME} sdev	0.000000	1.000000	1.000000
V_{IC} sdev	V_{MR} sdev	-0.003395	1.035088	0.945897
V_{ME} sdev	V_{MR} sdev	-0.003395	1.035088	0.945897

Shown in the first block are linear regressions of standard deviations from V_{EC}, V_{IC}, V_{ME}, and V_{MR} computed from the MCMC chain described in the legend for Table 2 with independent and dependent variables as indicated in the first two columns of the table. The second and third blocks are the same but for $\Delta = 3.0$ and $\Delta = 11.0$

Table 7. Regressions Among Covariances Demand and Supply Example

Variable				
Independent	Dependent	Intercept	Slope	R^2
		$\Delta = 0.9$		
V_{EC} sdev	V_{IC} sdev	-0.000021	1.018985	0.584151
V_{EC} sdev	V_{ME} sdev	-0.000075	1.640258	0.324489
V_{EC} sdev	V_{MR} sdev	-0.000081	2.483038	0.542015
V_{IC} sdev	V_{ME} sdev	-0.000044	1.975227	0.836412
V_{IC} sdev	V_{MR} sdev	-0.000031	2.457824	0.943967
V_{ME} sdev	V_{MR} sdev	0.000018	1.068931	0.832854
		$\Delta = 3.0$		
V_{EC} sdev	V_{IC} sdev	-0.000005	0.954483	0.509323
V_{EC} sdev	V_{ME} sdev	-0.000007	0.977459	0.498042
V_{EC} sdev	V_{MR} sdev	-0.000007	0.993356	0.497677
V_{IC} sdev	V_{ME} sdev	-0.000002	1.032686	0.994369
V_{IC} sdev	V_{MR} sdev	-0.000002	1.041013	0.977671
V_{ME} sdev	V_{MR} sdev	-0.000000	1.003528	0.974380
		$\Delta = 11.0$		
V_{EC} sdev	V_{IC} sdev	-0.000005	0.954483	0.509323
V_{EC} sdev	V_{ME} sdev	-0.000005	0.954483	0.509323
V_{EC} sdev	V_{MR} sdev	-0.000006	0.942636	0.484505
V_{IC} sdev	V_{ME} sdev	0.000000	1.000000	1.000000
V_{IC} sdev	V_{MR} sdev	-0.000001	1.001591	0.978439
V_{ME} sdev	V_{MR} sdev	-0.000001	1.001591	0.978439

Shown in the first block are linear regressions of covariances from V_{EC}, V_{IC}, V_{ME}, and V_{MR} computed from the MCMC chain described in the legend for Table 2 with independent and dependent variables as indicated in the first two columns of the table. The second and third blocks are the same but for $\Delta = 3.0$ and $\Delta = 11.0$

Table 8. Stochastic Discount Function Example, $\Delta = 2$

	Mean		Standard Deviation			
			Extrinsic		Modified	
Parameter	Extrinsic	Intrinsic	Extr Ctr	Intr Ctr	Extrinsic	Riemann
a_{01}	0.125950	0.130222	0.035219	0.035477	0.057211	0.085094
a_{02}	-0.008434	-0.016024	0.027108	0.028150	0.074924	0.053488
a_{03}	0.017429	0.013113	0.015427	0.016020	0.054562	0.031979
a_{04}	0.082601	0.075387	0.010530	0.012764	0.055936	0.024716
a_{05}	-0.061553	-0.074851	0.019684	0.023756	0.047917	0.043069
a_{06}	-0.036925	-0.024713	0.017924	0.021690	0.063208	0.043061
a_{07}	-0.028193	-0.010717	0.012460	0.021465	0.055240	0.040529
a_{08}	0.152953	0.164645	0.011347	0.016294	0.057191	0.033546
$b_{0,1}$	0.149272	0.159229	0.034798	0.036195	0.076915	0.072240
$b_{0,2}$	-0.246597	-0.268276	0.066741	0.070175	0.096793	0.154732
$B_{1,1}$	-0.046729	-0.034771	0.014092	0.018483	0.046557	0.027807
$B_{2,1}$	-0.058537	-0.036099	0.018411	0.029026	0.063679	0.006958
$B_{1,2}$	-0.007491	0.010769	0.019087	0.026416	0.075955	0.055285
$B_{2,2}$	-0.023266	-0.047909	0.023289	0.033908	0.082084	0.010080
$R_{0,1,1}$	0.836213	0.830120	0.026761	0.027446	0.074612	0.056217
$R_{0,1,2}$	-0.040340	-0.044094	0.010666	0.011308	0.049220	0.021801
$R_{0,2,2}$	0.993556	1.001678	0.042345	0.043117	0.108780	0.083124
$P_{1,1}$	0.551396	0.588314	0.052075	0.063836	0.102282	0.134788
$P_{2,2}$	0.099384	0.097378	0.053043	0.053081	0.106752	0.100766
a_1	-0.000000	-0.000008	0.000015	0.000016	0.000020	0.005618
a_2	-0.997967	-0.980331	0.010756	0.020659	0.043361	0.040626
a_3	-0.020725	0.013500	0.127623	0.132134	0.149616	0.272943

An MCMC chain of length 50,000 was computed using the Surface Sampling Algorithm for the SNP-ARCH likelihood (34) estimated from daily, inflation adjusted returns on the S&P500 and NASDAQ indices (including distributions) from January 1, 2010, to December 31, 2018 under moment conditions (2) as determined by (35) through (38). The prior for ρ is independent normal with location and scale the SNP-ARCH unconstrained maximum likelihood estimated parameters and standard errors. The prior for $\theta = (a_0, a_1, a_2)$ is independent normal with means $(0, -1, 0)$ and standard deviations $(1, 1, 1)$. The support conditions are normalizing sign restrictions on variance parameters and that the eigenvalues of the companion matrices for location and scale are less than one in absolute value. The chain was reduced with a stride of 10 leaving a chain of length 5,000 for computations. Means and standard deviations shown in the table for offset $\Delta = 2$, which is 1.0 larger than the smallest value of Δ for which the manifold M_ϵ is connected.

Table 9. Stochastic Discount Function Example, $\Delta = 10$

	Mean		Standard Deviation			
			Extrinsic		Modified	
Parameter	Extrinsic	Intrinsic	Extr Ctr	Intr Ctr	Extrinsic	Riemann
a_{01}	0.125950	0.118768	0.035219	0.035944	0.036786	0.036965
a_{02}	-0.008434	0.011021	0.027108	0.033367	0.034084	0.035347
a_{03}	0.017429	0.005592	0.015427	0.019446	0.020418	0.019750
a_{04}	0.082601	0.082544	0.010530	0.010530	0.011922	0.011040
a_{05}	-0.061553	-0.076661	0.019684	0.024814	0.026188	0.023665
a_{06}	-0.036925	-0.034718	0.017924	0.018060	0.019498	0.018794
a_{07}	-0.028193	-0.014846	0.012460	0.018261	0.019301	0.018503
a_{08}	0.152953	0.150794	0.011347	0.011551	0.012393	0.011898
$b_{0,1}$	0.149272	0.148000	0.034798	0.034821	0.037570	0.036086
$b_{0,2}$	-0.246597	-0.250884	0.066741	0.066879	0.067242	0.067979
$B_{1,1}$	-0.046729	-0.050522	0.014092	0.014594	0.015469	0.013500
$B_{2,1}$	-0.058537	-0.059758	0.018411	0.018451	0.019506	0.002549
$B_{1,2}$	-0.007491	-0.020881	0.019087	0.023316	0.024502	0.023297
$B_{2,2}$	-0.023266	-0.041644	0.023289	0.029668	0.032058	0.004874
$R_{0,1,1}$	0.836213	0.827857	0.026761	0.028036	0.030673	0.029207
$R_{0,1,2}$	-0.040340	-0.028008	0.010666	0.016305	0.016786	0.019185
$R_{0,2,2}$	0.993556	0.960481	0.042345	0.053734	0.054459	0.054449
$P_{1,1}$	0.551396	0.539821	0.052075	0.053346	0.054371	0.055694
$P_{2,2}$	0.099384	0.101043	0.053043	0.053069	0.055329	0.054195
a_1	-0.000000	-0.000002	0.000015	0.000015	0.000015	0.003165
a_2	-0.997967	-0.998806	0.010756	0.010789	0.011167	0.011660
a_3	-0.020725	0.008479	0.127623	0.130923	0.130910	0.139620

As for Table 8 except that $\Delta = 10$, which is the point just after the curves in Figure 4 begin to flatten.

Table 10. Stochastic Discount Function Example, $\Delta = 31$

	Mean		Standard Deviation			
			Extrinsic		Modified	
Parameter	Extrinsic	Intrinsic	Extr Ctr	Intr Ctr	Extrinsic	Riemann
a_{01}	0.125950	0.118768	0.035219	0.035944	0.035940	0.036115
a_{02}	-0.008434	0.011021	0.027108	0.033367	0.033364	0.033468
a_{03}	0.017429	0.005592	0.015427	0.019446	0.019444	0.019369
a_{04}	0.082601	0.082544	0.010530	0.010530	0.010529	0.010539
a_{05}	-0.061553	-0.076661	0.019684	0.024814	0.024812	0.025392
a_{06}	-0.036925	-0.034718	0.017924	0.018060	0.018058	0.018097
a_{07}	-0.028193	-0.014846	0.012460	0.018261	0.018259	0.018050
a_{08}	0.152953	0.150794	0.011347	0.011551	0.011550	0.011555
$b_{0,1}$	0.149272	0.148000	0.034798	0.034821	0.034818	0.034725
$b_{0,2}$	-0.246597	-0.250884	0.066741	0.066879	0.066872	0.066737
$B_{1,1}$	-0.046729	-0.050522	0.014092	0.014594	0.014592	0.014727
$B_{2,1}$	-0.058537	-0.059758	0.018411	0.018451	0.018449	0.018453
$B_{1,2}$	-0.007491	-0.020881	0.019087	0.023316	0.023314	0.023764
$B_{2,2}$	-0.023266	-0.041644	0.023289	0.029668	0.029665	0.029748
$R_{0,1,1}$	0.836213	0.827857	0.026761	0.028036	0.028033	0.028009
$R_{0,1,2}$	-0.040340	-0.028008	0.010666	0.016305	0.016304	0.016159
$R_{0,2,2}$	0.993556	0.960481	0.042345	0.053734	0.053728	0.053690
$P_{1,1}$	0.551396	0.539821	0.052075	0.053346	0.053341	0.053352
$P_{2,2}$	0.099384	0.101043	0.053043	0.053069	0.053064	0.053072
a_1	-0.000000	-0.000002	0.000015	0.000015	0.000015	0.000013
a_2	-0.997967	-0.998806	0.010756	0.010789	0.010788	0.008667
a_3	-0.020725	0.008479	0.127623	0.130923	0.130910	0.130925

As for Table 8 except that $\Delta = 31$, which is the smallest value such that each node in M_ϵ is connected to all other nodes.

Table 11. Regressions Among Standard Deviations, Stochastic Discount Function Example

Variable				
Independent	Dependent	Intercept	Slope	R^2
		$\Delta = 2.0$		
V_{EC} sdev	V_{IC} sdev	0.004282	0.996627	0.979527
V_{EC} sdev	V_{ME} sdev	0.041461	0.971254	0.769701
V_{EC} sdev	V_{MR} sdev	0.000142	2.148869	0.941703
V_{IC} sdev	V_{ME} sdev	0.037353	0.972626	0.782703
V_{IC} sdev	V_{MR} sdev	-0.007529	2.109841	0.920539
V_{ME} sdev	V_{MR} sdev	-0.050916	1.632486	0.666095
		$\Delta = 10.0$		
V_{EC} sdev	V_{IC} sdev	0.002420	1.005033	0.987339
V_{EC} sdev	V_{ME} sdev	0.003663	1.001094	0.986990
V_{EC} sdev	V_{MR} sdev	-0.000386	1.077904	0.949659
V_{IC} sdev	V_{ME} sdev	0.001260	0.995870	0.999223
V_{IC} sdev	V_{MR} sdev	-0.002833	1.067869	0.953538
V_{ME} sdev	V_{MR} sdev	-0.004076	1.069052	0.948511
		$\Delta = 31.0$		
V_{EC} sdev	V_{IC} sdev	0.002420	1.005033	0.987339
V_{EC} sdev	V_{ME} sdev	0.002420	1.005033	0.987339
V_{EC} sdev	V_{MR} sdev	0.002308	1.006940	0.986297
V_{IC} sdev	V_{ME} sdev	-0.000000	1.000000	1.000000
V_{IC} sdev	V_{MR} sdev	-0.000129	1.002270	0.999689
V_{ME} sdev	V_{MR} sdev	-0.000129	1.002270	0.999689

Shown in the first block are linear regressions of standard deviations from V_{EC}, V_{IC}, V_{ME}, and V_{MR} computed from the MCMC chain described in the legend for Table 8 with independent and dependent variables as indicated in the first two columns of the table. The second and third blocks are the same but for $\Delta = 10.0$ and $\Delta = 31.0$

Table 12. Regressions Among Covariances, Stochastic Discount Function Example

Variable				
Independent	Dependent	Intercept	Slope	R^2
		$\Delta = 2.0$		
V_{EC} sdev	V_{IC} sdev	0.000017	1.103377	0.698067
V_{EC} sdev	V_{ME} sdev	0.000071	2.383066	0.242961
V_{EC} sdev	V_{MR} sdev	0.000110	3.960890	0.457034
V_{IC} sdev	V_{ME} sdev	0.000035	3.060979	0.699095
V_{IC} sdev	V_{MR} sdev	0.000049	3.714154	0.700864
V_{ME} sdev	V_{MR} sdev	0.000020	0.850015	0.491984
		$\Delta = 10.0$		
V_{EC} sdev	V_{IC} sdev	-0.000002	1.013226	0.830676
V_{EC} sdev	V_{ME} sdev	-0.000004	1.066309	0.819366
V_{EC} sdev	V_{MR} sdev	-0.000010	1.050009	0.842409
V_{IC} sdev	V_{ME} sdev	-0.000002	1.055850	0.992882
V_{IC} sdev	V_{MR} sdev	-0.000009	0.981551	0.909796
V_{ME} sdev	V_{MR} sdev	-0.000007	0.917860	0.893262
		$\Delta = 31.0$		
V_{EC} sdev	V_{IC} sdev	-0.000002	1.013226	0.830676
V_{EC} sdev	V_{ME} sdev	-0.000002	1.013226	0.830676
V_{EC} sdev	V_{MR} sdev	-0.000002	1.014812	0.831122
V_{IC} sdev	V_{ME} sdev	0.000000	1.000000	1.000000
V_{IC} sdev	V_{MR} sdev	-0.000000	1.001103	0.999615
V_{ME} sdev	V_{MR} sdev	-0.000000	1.001103	0.999615

Shown in the first block are linear regressions of covariances from V_{EC}, V_{IC}, V_{ME}, and V_{MR} computed from the MCMC chain described in the legend for Table 8 with independent and dependent variables as indicated in the first two columns of the table. The second and third blocks are the same but for $\Delta = 10.0$ and $\Delta = 31.0$

Table 13. Curved Manifold Example, $\Delta = 0.57$

	Mean		Standard Deviation or Correlation			
			Extrinsic		Modified	
Parameter	Extrinsic	Intrinsic	Extr Ctr	Intr Ctr	Extrinsic	Riemann
μ_1	0.003030	0.001782	0.044938	0.044956	0.256304	0.045930
μ_2	0.010777	0.008102	0.046710	0.046787	0.282870	0.047894
$R_{1,1}$	0.997487	0.992473	0.030209	0.030622	0.199155	0.031385
$R_{1,2}$	-0.011216	-0.008383	0.021103	0.021293	0.133476	0.021763
$R_{2,2}$	1.029374	1.030792	0.010518	0.010614	0.066102	0.010639
θ	5.379109	5.377738	0.155378	0.155384	0.975752	0.159227
$\rho(\mu_1, \mu_2)$			-0.078107	-0.076362	-0.043754	-0.075747
$\rho(\mu_1, R_{1,1})$			-0.038925	-0.033837	-0.031125	-0.030951
$\rho(\mu_1, R_{1,2})$			-0.014263	-0.017826	-0.008660	-0.010185
$\rho(\mu_1, R_{2,2})$			-0.049502	-0.052750	-0.032077	-0.050698
$\rho(\mu_1, \theta)$			-0.034030	-0.033771	-0.025612	-0.031390
$\rho(\mu_2, R_{1,1})$			-0.000003	0.009360	-0.028539	0.010071
$\rho(\mu_2, R_{1,2})$			0.061739	0.053481	-0.018607	0.054120
$\rho(\mu_2, R_{2,2})$			-0.230838	-0.236033	-0.068083	-0.223426
$\rho(\mu_2, \theta)$			0.003121	0.003620	-0.025818	0.003372
$\rho(R_{1,1}, R_{1,2})$			-0.149040	-0.167514	0.333217	-0.161162
$\rho(R_{1,1}, R_{2,2})$			0.439249	0.407558	0.463633	0.440073
$\rho(R_{1,1}, \theta)$			0.467925	0.463035	0.475693	0.462643
$\rho(R_{1,2}, R_{2,2})$			0.762028	0.766259	0.793649	0.771083
$\rho(R_{1,2}, \theta)$			0.801051	0.792721	0.819296	0.798896
$\rho(R_{2,2}, \theta)$			0.960119	0.950296	0.936706	0.973192

The data are a simulation of the curved manifold example. An MCMC chain of length 50,000 was computed using the Surface Sampling Algorithm for the normal likelihood (39) subject to moment conditions (2) as determined by (41) and (42) The prior for ρ is independent normal with location the unconstrained maximum likelihood estimates of (39) and scale 5.0. The prior for θ is normal with mean 5.0 and standard deviations 5.0. The support conditions on R are that diagonals must be positive and θ must be positive.. The chain was reduced by eliminating repetitions due to rejections to a length of 37,269 for computations. Means and standard deviations shown in the table for offset $\Delta = 0.57$, which is the smallest value of Δ for which the manifold M_ϵ is connected.

Table 14. Curved Manifold Example, $\Delta = 3.0$

	Mean		Standard Deviation or Correlation			
			Extrinsic		Modified	
Parameter	Extrinsic	Intrinsic	Extr Ctr	Intr Ctr	Extrinsic	Riemann
μ_1	0.003030	0.001782	0.044938	0.044956	0.064382	0.044967
μ_2	0.010777	0.008102	0.046710	0.046787	0.069694	0.046818
$R_{1,1}$	0.997487	0.992473	0.030209	0.030622	0.049044	0.030692
$R_{1,2}$	-0.011216	-0.008383	0.021103	0.021293	0.034104	0.021243
$R_{2,2}$	1.029374	1.030792	0.010518	0.010614	0.015268	0.010386
θ	5.379109	5.377738	0.155378	0.155384	0.228527	0.155419
$\rho(\mu_1, \mu_2)$			-0.078107	-0.076362	-0.039940	-0.076271
$\rho(\mu_1, R_{1,1})$			-0.038925	-0.033837	-0.013112	-0.031901
$\rho(\mu_1, R_{1,2})$			-0.014263	-0.017826	0.006544	-0.012268
$\rho(\mu_1, R_{2,2})$			-0.049502	-0.052750	-0.003076	-0.053070
$\rho(\mu_1, \theta)$			-0.034030	-0.033771	-0.004220	-0.033868
$\rho(\mu_2, R_{1,1})$			-0.000003	0.009360	0.000939	0.009883
$\rho(\mu_2, R_{1,2})$			0.061739	0.053481	0.020551	0.054497
$\rho(\mu_2, R_{2,2})$			-0.230838	-0.236033	-0.046810	-0.223549
$\rho(\mu_2, \theta)$			0.003121	0.003620	0.013745	0.003537
$\rho(R_{1,1}, R_{1,2})$			-0.149040	-0.167514	0.163611	-0.162194
$\rho(R_{1,1}, R_{2,2})$			0.439249	0.407558	0.320706	0.440313
$\rho(R_{1,1}, \theta)$			0.467925	0.463035	0.329882	0.462884
$\rho(R_{1,2}, R_{2,2})$			0.762028	0.766259	0.653799	0.770204
$\rho(R_{1,2}, \theta)$			0.801051	0.792721	0.698690	0.798101
$\rho(R_{2,2}, \theta)$			0.960119	0.950296	0.886439	0.973122

As for Table 13 except that $\Delta = 3.0$, which is the point just after the curves in Figure 4 begin to flatten.

Table 15. Curved Manifold Example, $\Delta = 15.0$

	Mean		Standard Deviation or Correlation			
			Extrinsic		Modified	
Parameter	Extrinsic	Intrinsic	Extr Ctr	Intr Ctr	Extrinsic	Riemann
μ_1	0.003030	0.001782	0.044938	0.044956	0.045979	0.044963
μ_2	0.010777	0.008102	0.046710	0.046787	0.049665	0.046814
$R_{1,1}$	0.997487	0.992473	0.030209	0.030622	0.036357	0.030688
$R_{1,2}$	-0.011216	-0.008383	0.021103	0.021293	0.024844	0.021239
$R_{2,2}$	1.029374	1.030792	0.010518	0.010614	0.010586	0.010383
θ	5.379109	5.377738	0.155378	0.155384	0.158211	0.155382
$\rho(\mu_1, \mu_2)$			-0.078107	-0.076362	-0.017182	-0.076274
$\rho(\mu_1, R_{1,1})$			-0.038925	-0.033837	-0.003856	-0.031906
$\rho(\mu_1, R_{1,2})$			-0.014263	-0.017826	-0.022850	-0.012280
$\rho(\mu_1, R_{2,2})$			-0.049502	-0.052750	-0.035418	-0.053089
$\rho(\mu_1, \theta)$			-0.034030	-0.033771	-0.033630	-0.033885
$\rho(\mu_2, R_{1,1})$			-0.000003	0.009360	0.011497	0.009885
$\rho(\mu_2, R_{1,2})$			0.061739	0.053481	0.028671	0.054493
$\rho(\mu_2, R_{2,2})$			-0.230838	-0.236033	-0.083846	-0.223585
$\rho(\mu_2, \theta)$			0.003121	0.003620	0.016311	0.003531
$\rho(R_{1,1}, R_{1,2})$			-0.149040	-0.167514	-0.047995	-0.162304
$\rho(R_{1,1}, R_{2,2})$			0.439249	0.407558	0.186387	0.440271
$\rho(R_{1,1}, \theta)$			0.467925	0.463035	0.209996	0.462847
$\rho(R_{1,2}, R_{2,2})$			0.762028	0.766259	0.448719	0.770156
$\rho(R_{1,2}, \theta)$			0.801051	0.792721	0.519493	0.798059
$\rho(R_{2,2}, \theta)$			0.960119	0.950296	0.797676	0.973115

As for Table 13 except that $\Delta = 15.0$, which is the smallest value of Δ such that each node in M_ϵ is connected to all other nodes.

Table 16. Regressions Among Standard Deviations, Curved Manifold Example

Variable				
Independent	Dependent	Intercept	Slope	R^2
		$\Delta = 0.57$		
V_{EC} sdev	V_{IC} sdev	0.000201	0.998685	0.999993
V_{EC} sdev	V_{ME} sdev	-0.004274	6.279022	0.998619
V_{EC} sdev	V_{MR} sdev	0.000080	1.024291	0.999985
V_{IC} sdev	V_{ME} sdev	-0.005557	6.287697	0.998756
V_{IC} sdev	V_{MR} sdev	-0.000126	1.025643	0.999998
V_{ME} sdev	V_{MR} sdev	0.000844	0.162919	0.998793
		$\Delta = 3.0$		
V_{EC} sdev	V_{IC} sdev	0.000201	0.998685	0.999993
V_{EC} sdev	V_{ME} sdev	0.001779	1.458109	0.999163
V_{EC} sdev	V_{MR} sdev	0.000132	0.999614	0.999985
V_{IC} sdev	V_{ME} sdev	0.001482	1.460118	0.999292
V_{IC} sdev	V_{MR} sdev	-0.000069	1.000932	0.999997
V_{ME} sdev	V_{MR} sdev	-0.001049	0.685046	0.999337
		$\Delta = 15.0$		
V_{EC} sdev	V_{IC} sdev	0.000201	0.998685	0.999993
V_{EC} sdev	V_{ME} sdev	0.002700	1.001903	0.998382
V_{EC} sdev	V_{MR} sdev	0.000135	0.999362	0.999985
V_{IC} sdev	V_{ME} sdev	0.002494	1.003311	0.998565
V_{IC} sdev	V_{MR} sdev	-0.000066	1.000680	0.999996
V_{ME} sdev	V_{MR} sdev	-0.002478	0.995999	0.998667

Shown in the first block are linear regressions of standard deviations from V_{EC}, V_{IC}, V_{ME}, and V_{MR} computed from the MCMC chain described in the legend for Table 13 with independent and dependent variables as indicated in the first two columns of the table. The second and third blocks are the same but for $\Delta = 3.0$ and $\Delta = 15.0$

Table 17. Regressions Among Covariances, Curved Manifold Example

Variable				
Independent	Dependent	Intercept	Slope	R^2
		$\Delta = 0.57$		
V_{EC} sdev	V_{IC} sdev	-0.000000	1.000302	0.999944
V_{EC} sdev	V_{ME} sdev	0.000985	40.107354	0.985408
V_{EC} sdev	V_{MR} sdev	0.000004	1.050841	0.999935
V_{IC} sdev	V_{ME} sdev	0.001009	40.078723	0.984650
V_{IC} sdev	V_{MR} sdev	0.000004	1.050508	0.999961
V_{ME} sdev	V_{MR} sdev	-0.000015	0.025815	0.985080
		$\Delta = 3.0$		
V_{EC} sdev	V_{IC} sdev	-0.000000	1.000302	0.999944
V_{EC} sdev	V_{ME} sdev	0.000114	1.861595	0.982616
V_{EC} sdev	V_{MR} sdev	0.000002	1.002748	0.999959
V_{IC} sdev	V_{ME} sdev	0.000115	1.860169	0.981758
V_{IC} sdev	V_{MR} sdev	0.000002	1.002431	0.999986
V_{ME} sdev	V_{MR} sdev	-0.000052	0.529139	0.982031
		$\Delta = 15.0$		
V_{EC} sdev	V_{IC} sdev	-0.000000	1.000302	0.999944
V_{EC} sdev	V_{ME} sdev	0.000016	0.705615	0.962668
V_{EC} sdev	V_{MR} sdev	0.000002	1.002291	0.999958
V_{IC} sdev	V_{ME} sdev	0.000017	0.705142	0.962014
V_{IC} sdev	V_{MR} sdev	0.000002	1.001975	0.999986
V_{ME} sdev	V_{MR} sdev	-0.000005	1.367140	0.962229

Shown in the first block are linear regressions of covariances from V_{EC}, V_{IC}, V_{ME}, and V_{MR} computed from the MCMC chain described in the legend for Table 13 with independent and dependent variables as indicated in the first two columns of the table. The second and third blocks are the same but for $\Delta = 3.0$ and $\Delta = 15.0$