Stochastic Input Modeling considering Dependencies: Generating Multivariate Time Series using a Vector Autoregressive Technique*

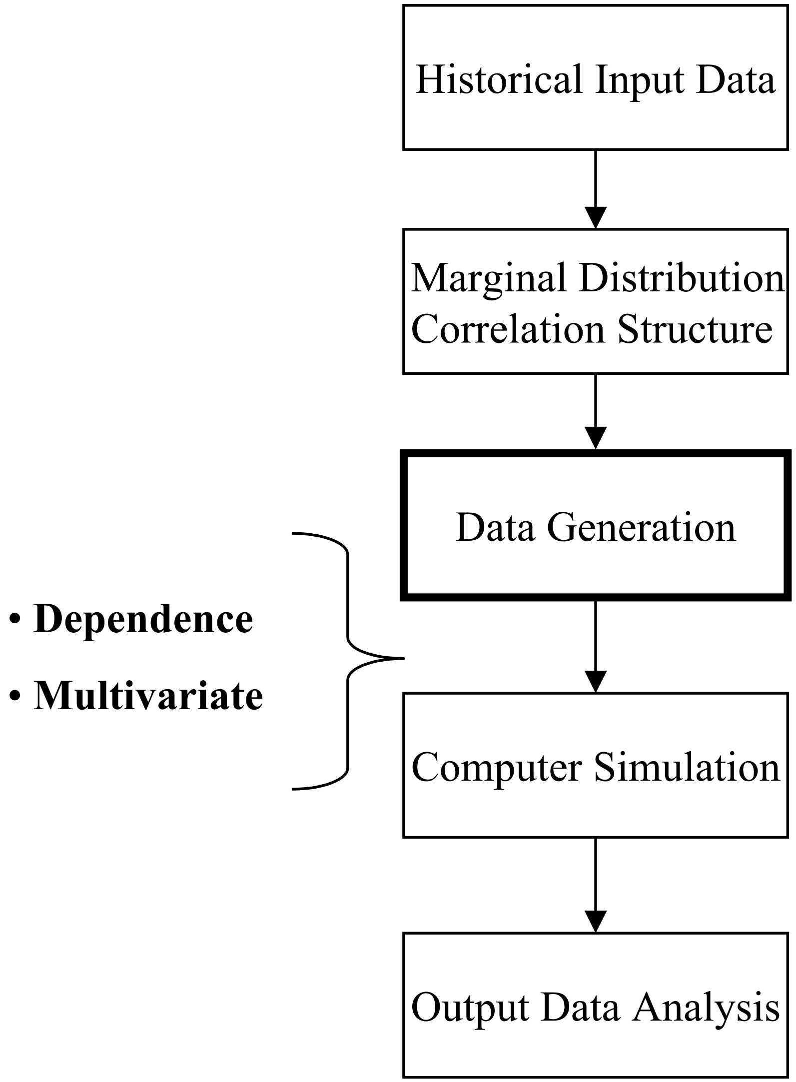
Bahar Deler & Barry L. Nelson

Northwestern University

November 8, 2000

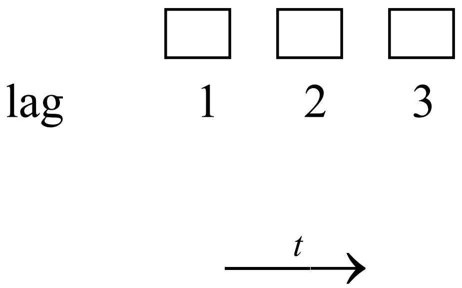
*Research supported by NSF grant DMI 9821011

Stochastic Input Modeling



Problem Setup & Contribution

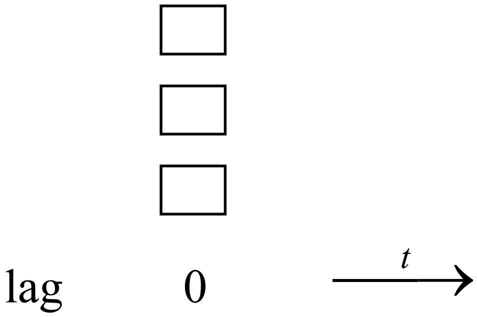
Univariate Time Series



Approaches:

- Exploit properties specific to the marginal distribution (Lewis et al. 1989)
- Construct a series of autocorrelated uniform random variables as the base process and apply the inverse transform method (Melamed 1991, Cario and Nelson 1996)

Random Vectors



Approach:

- Transform multivariate copula into vectors with arbitrary marginal distributions (Cario et al. 2000, Chen 2000, Ghosh and Henderson 2000)

Develop a comprehensive framework that pulls together the theory behind univariate time series and random vectors with dependent components and extend it to the multivariate time series

Modeling Approach

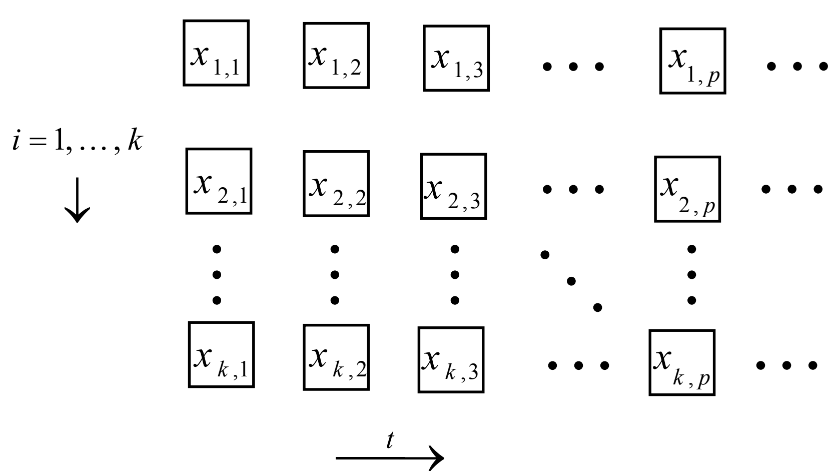
Customize a flexible model incorporating

- a vector auto-regressive process
- Johnson family of marginal distributions

<u>Step 0</u>

- For the input process X_t, obtain Johnson marginal distributions for each of the k series specified through their first four moments and the correlation matrices for lags ranging from 0 to p

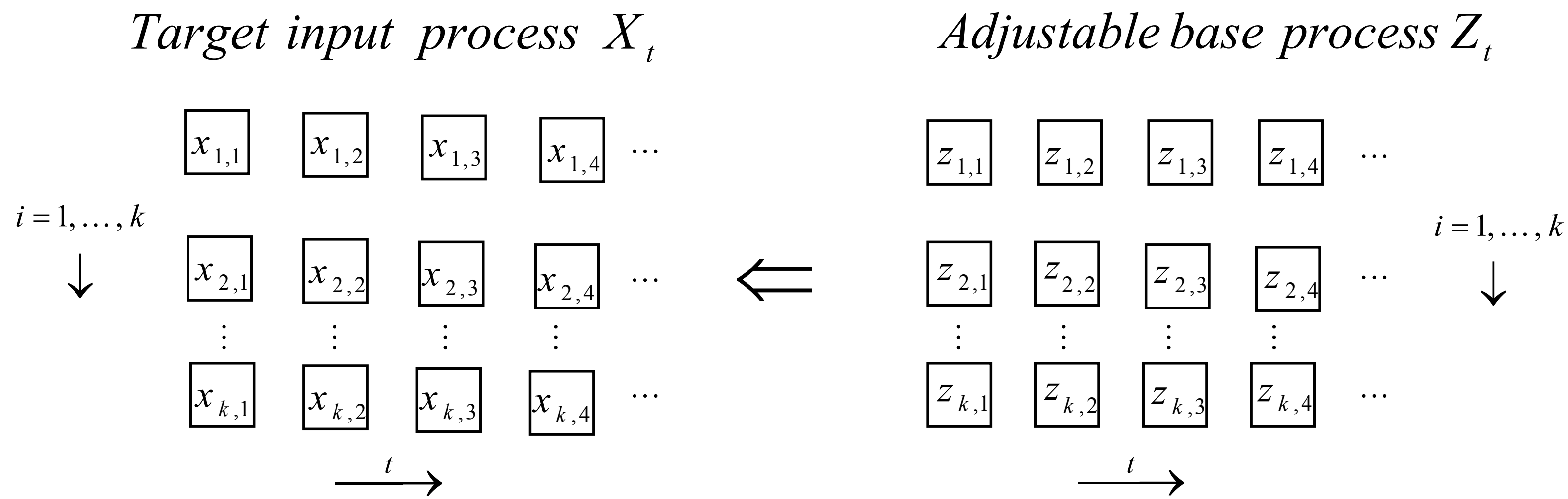
Input process X_t



Modeling Approach (cont'd)

Step 1

– Characterize the underlying vector autoregressive model , Z_t, by solving $pk^2 + k(k-1)/2$ correlation matching problems



– Determine system parameters using the correlation matrices $\Sigma_Z(0), \ldots, \Sigma_Z(p)$

– Generate standard normal random variates, $z_{i,t}$, $i = 1, \ldots, k$, $t = 0, \ldots, T$, using the system parameters

Step 2

– Transform $z_{i,t}$ into Johnson variates $x_{i,t}$ by $x_{i,t} = \varepsilon_i + \lambda_i \, g_i^{-1}\left(\frac{z_{i,t} - \gamma_i}{\delta_i}\right)$

Johnson Family of Distributions

Definition: For a Johnson variable X, $F_X(x) = \Phi\left\{\gamma + \delta\, g\left(\frac{x-\varepsilon}{\lambda}\right)\right\}$

$$g(y) = \begin{cases} y & \textit{for } S_N \textit{ (normal) family,} \\ \log(y) & \textit{for } S_L \textit{ (lognormal) family,} \\ \sinh^{-1}(y) & \textit{for } S_U \textit{ (unbounded) family,} \\ \log\left(\frac{y}{1-y}\right) & \textit{for } S_B \textit{ (bounded) family.} \end{cases}$$

Φ : standard normal cumulative distribution function

$(\gamma, \delta, \lambda, \varepsilon)$: Johnson parameters

Characteristics

- Match any feasible first four moments
- Provide unique distribution for each combination of skewness & kurtosis
- Generate random variates: $x = \varepsilon + \lambda\, g^{-1}\left(\frac{z-\gamma}{\delta}\right)$ where $z \approx N(0,1)$

Vector Autoregressive Model

Definition

$$Z_t = \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + \ldots + \alpha_p Z_{t-p} + u_t, \ t = 0, \pm 1, \pm 2, \ldots$$

– Z_t is $(k \, x \, 1)$, α_i is $(k \, x \, k)$, u_t is $(k \, x \, 1)$

Assumption

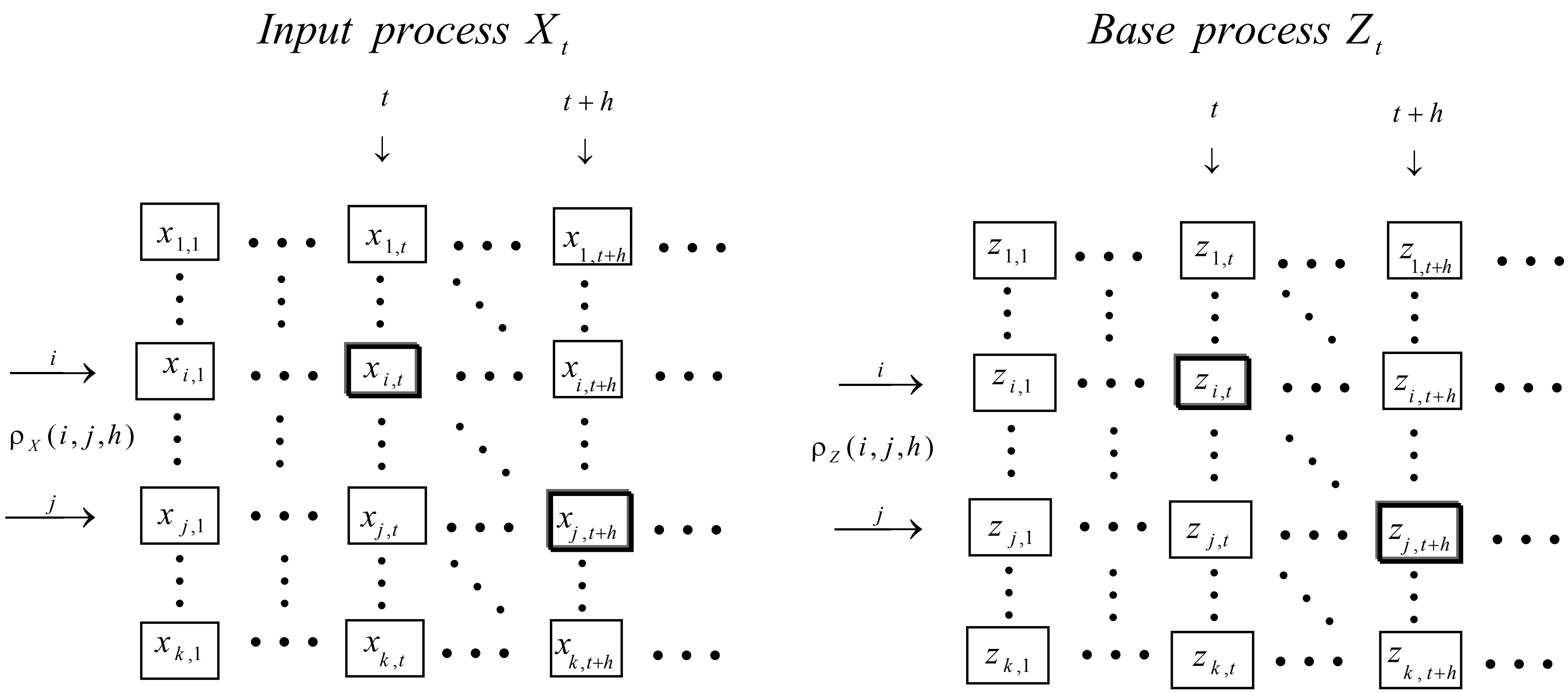
– Z_t is stationarity $\Rightarrow$ stationarity of the input process X_t
– $u_t \approx N_k(0, \Sigma_u)$ $\Rightarrow$ Z_t is a Gaussian process

System Parameter Estimation

– Determine $\alpha_1, \ldots, \alpha_p$ using multivariate Yule-Walker equations
– Solve $\Sigma_u = \Sigma_Z(0) - \alpha_1 \Sigma_Z^{'}(1) - \ldots - \alpha_p \Sigma_Z^{'}(p)$

Correlation Matching Problem

<u>Objective</u>: Select the covariance structure for the vector autoregressive process that gives the desired covariance structure for the input process



$$c_{ijh}\left[\rho_Z(i,j,h)\right] = \rho_X(i,j,h)$$ **(independent matching problems)**

Correlation Matching Problem

$$Corr\left[X_{i,t}, X_{j,t+h}\right] = \rho_X(i,j,h) = \frac{E\left[X_{i,t} * X_{j,t+h}\right] - \mu_i \mu_j}{\sigma_i \quad \sigma_j}$$

$$E\left[X_{i,t} X_{j,t+h}\right] = E\left\{F_{X_i}^{-1}\left[\Phi\left(Z_{i,t}\right)\right] * F_{X_i}^{-1}\left[\Phi\left(Z_{j,t+h}\right)\right]\right\}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \underbrace{F_{X_i}^{-1}\left[\Phi\left(z_i\right)\right]}_{x_i = \varepsilon_i + \lambda_i g_i^{-1}\left(\frac{z_i - \gamma_i}{\delta_i}\right)} \underbrace{F_{X_j}^{-1}\left[\Phi\left(z_j\right)\right]}_{x_j = \varepsilon_j + \lambda_j g_j^{-1}\left(\frac{z_j - \gamma_j}{\delta_j}\right)} \overbrace{\vartheta\left(z_i, z_j, \rho_Z(i,j,h)\right)}^{\rho_Z(i,j,h) \text{ appears only here}} dz_i \, dz_j$$

Rewriting the correlation function as

$$c_{ijh}\left[\rho_Z(i,j,h)\right] = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\varepsilon_i + \lambda_i f_i^{-1}\left(\frac{z_i - \gamma_i}{\delta_i}\right)\right]\left[\varepsilon_j + \lambda_j f_j^{-1}\left(\frac{z_j - \gamma_j}{\delta_j}\right)\right] \vartheta\left(z_i, z_j; \rho_Z(i,j,h)\right) dz_i \, dz_j - \mu_i \mu_j}{\sigma_i \sigma_j}$$

$$\Rightarrow \quad \begin{matrix} i,j = 1,\ldots,k \\ i \neq j \\ h = 0,\ldots,p \end{matrix} \quad c_{ijh}\left[\rho_Z(i,j,h)\right] = \rho_X(i,j,h) \quad \Rightarrow \quad pk^2 + k(k-1)/2 \ \text{ matching problems}$$

Implementation

Objective: Establish properties of the function $c_{ijh}\left[\rho_Z(i,j,h)\right]$ to perform an efficient numerical search to find $\rho_Z(i,j,h)$ within a predetermined precision

Useful Results

- $c_{ijh}\left[0\right]=0$
- $\rho_Z(i,j,h) \geq (\leq) 0 \;\Rightarrow\; \rho_X(i,j,h) \geq (\leq) 0$
- Extremal values of $\rho_X(i,j,h)$ are attainable
- $c_{ijh}\left[\rho_Z(i,j,h)\right]$ is continuous and non-decreasing for $-1 \leq \rho_Z(i,j,h) \leq 1$

Any feasible bivariate correlation for F_{X_i} and F_{X_j} is attainable

Implementation (cont'd)

- The regula falsi (secant) solves each correlation matching problem very accurately and efficiently, particularly for the Johnson system:
  - Define $w[\rho_Z] = c[\rho_Z] - \rho_X$
  - If $\rho_X > 0 \Rightarrow y_0 = 0 \ (w(0) < 0), \ y_1 = 1 \ (w(1) > 0)$

    Else $\rho_X < 0 \Rightarrow y_0 = 0 \ (w(0) < 0), \ y_1 = -1 \ (w(-1) < 0)$
  - Iterative procedure is given by $y_{m+1} = y_m - w(y_m) \dfrac{y_m - y_{m-1}}{w(y_m) - w(y_{m-1})}$
  - Convergence is certain as $w(x_0)$ and $w(x_1)$ have opposite sign.
- Evaluate $w(y_m)$ through a numerical integration using a doubly-infinite hyper-cube transformation
- Estimate an accurate and reliable error for $w(y_m)$ using null rules

Hyper-cube Transformation

$$\psi(z_i^*) = \tan\left(\pi z_i^* \big/ 2\right), \; -1 < z_i^* < 1, \; \partial\psi(z_i^*) \big/ \partial z_i^* = \pi \big/ 2 \left(1 + \tan^2\left(\pi z_i^* \big/ 2\right)\right), \; i = 1, 2.$$

$\Rightarrow$

For $|\rho| < 1:$

$$\int_{-1}^{1}\int_{-1}^{1} \frac{w(\tan(\pi z_1^*/2), \tan(\pi z_2^*/2)) \quad g(\tan(\pi z_1^*/2), \tan(\pi z_2^*/2), \rho)}{4/\pi^2 (1+\tan^2(\pi z_1^*/2))^{-1} \quad (1+\tan^2(\pi z_2^*/2))^{-1}} \, dz_1 \, dz_2$$

where $w(z_1, z_2) = e^{-(z_1^2 + z_2^2)}$

For $\rho = 1$ *and* $\rho = -1:$

$$\begin{cases} \displaystyle\int_{-1}^{1}\int_{-1}^{1} \frac{[\varepsilon_1 + \lambda_1 f_1^{-1}((\tan(\pi z_1^*/2) - \gamma_1)/\delta_1)] \quad [\varepsilon_2 + \lambda_2 f_2^{-1}((\tan(\pi z_2^*/2) - \gamma_{21})/\delta_{21})]}{2\sqrt{2}\sqrt{\pi^{-1}} \quad (1+\tan^2(\pi z_1^*/2))^{-1} \quad e^{\tan^2(\pi z_2^*/2)/2}} dz_1^* \, dz_2^* \\ \displaystyle\int_{-1}^{1}\int_{-1}^{1} \frac{[\varepsilon_1 + \lambda_1 f_1^{-1}((\tan(\pi z_1^*/2)\gamma_1)/\delta_1)] \quad [\varepsilon_2 + \lambda_2 f_2^{-1}((-\tan(\pi z_2^*/2)\gamma_{21})/\delta_{21})]}{2\sqrt{2}\sqrt{\pi^{-1}} \quad (1+\tan^2(\pi z_1^*/2))^{-1} \quad e^{\tan^2(\pi z_2^*/2)/2}} dz_1^* \, dz_2^* \end{cases}$$

Null Rules

Definition $$N_d(\ell) = \sum_{i=1}^{n} u_i \, \ell\left(x_i\right), \; \exists u_i \neq 0, \; \sum_{i=1}^{n} u_i = 0.$$

Characteristics

- Maps to zero all polynomials of degree not more than d
- Equally strong with the cubature formula $\sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} c_i^2$
- Decrease the probability of phase factor effects
- Heuristically distinguish possible types of behavior to extract an error estimate:

$$r^2 = \max\left\{ \frac{N_3(\ell)}{N_1(\ell)}, \frac{N_5(\ell)}{N_3(\ell)}, \frac{N_7(\ell)}{N_5(\ell)} \right\}$$

$$\Rightarrow \quad \begin{aligned} & \textit{If } \; r < 1, \textit{ then strong asymptotic behavior} \\ & \qquad \Rightarrow \quad \textit{error estimate} \cong a \, r^{q-s} N_s(\ell), \\ & \textit{Else if } \; r_{crit} \le r \le 1, \textit{ then weak asymptotic behavior} \\ & \qquad \Rightarrow \quad \textit{error estimate} \cong K \, r^2 N_s(\ell), \quad a = K \, r_{crit}^{s-q+2}. \\ & \textit{Else if } \; r > 1, \textit{ then no asymptotic behavior} \\ & \qquad \Rightarrow \quad \textit{error estimate} \cong K \, N_s(\ell). \end{aligned}$$

Example

Step 0

A bivariate random variable with marginals and correlation given as

– Unbounded distribution specified by $\mu = 0, \sigma = 1, \beta_1 = 0.5, \beta_2 = 5$

$$\Rightarrow \gamma_1 = 0, \delta_1 = 1.875, \lambda_1 = 1.875, \varepsilon_1 = 0$$

– Bounded distribution specified by $\mu = 0, \sigma = 1, \beta_1 = 1, \beta_2 = 2.8$

$$\Rightarrow \gamma_2 = 0.478, \delta_2 = 0.375, \lambda_2 = 4.123, \varepsilon_2 = -0.859$$

– Correlation matrices are given at lag 0 and lag 1 ($k = 2, p = 1$)

$$\sum_X(0) = \begin{pmatrix} 1 & 0.801 \\ 0.801 & 1 \end{pmatrix} \quad and \quad \sum_X(1) = \begin{pmatrix} 0.829 & 0.309 \\ 0.516 & 0.713 \end{pmatrix}$$

Example (cont'd)

Step 1

- Characterize the underlying process by solving 5 correlation matching problems

$$\sum_Z(0)=\begin{pmatrix} 1 & 0.599 \\ 0.599 & 1 \end{pmatrix} \quad and \quad \sum_Z(1)=\begin{pmatrix} 0.621 & 0.232 \\ 0.388 & 0.534 \end{pmatrix}$$

- Using the correlation structure, determine the system parameters for

$$Z_t = \alpha_1 Z_{t-1} + u_t \,,\; t = 0, 1, \ldots$$

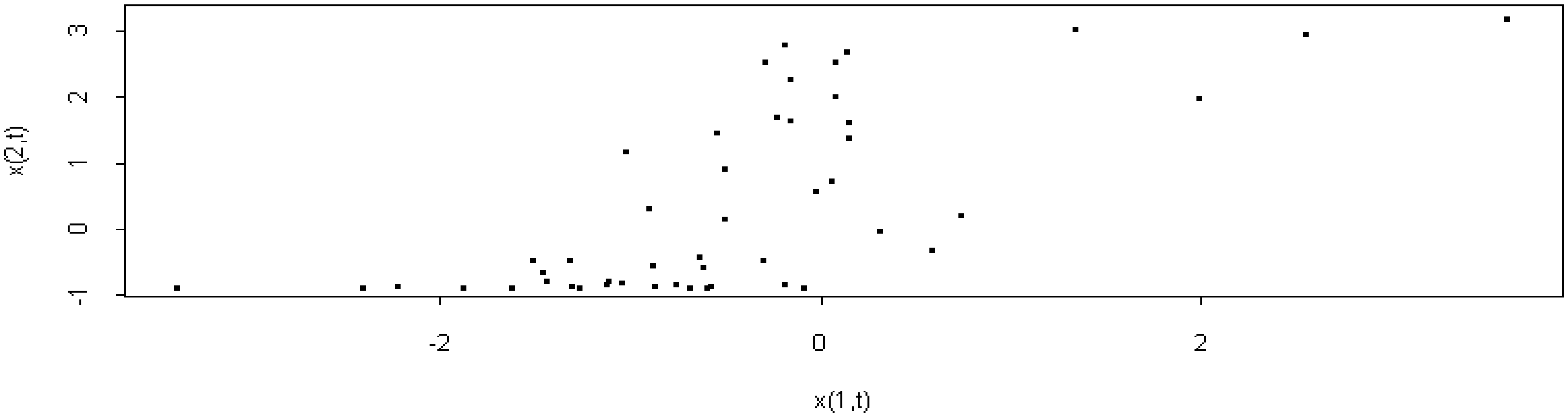
$$\alpha_1 = \begin{pmatrix} 0.752 & -0.218 \\ 0.105 & 0.471 \end{pmatrix} \qquad \sum_u = \begin{pmatrix} 0.584 & 0.425 \\ 0.425 & 0.708 \end{pmatrix}$$

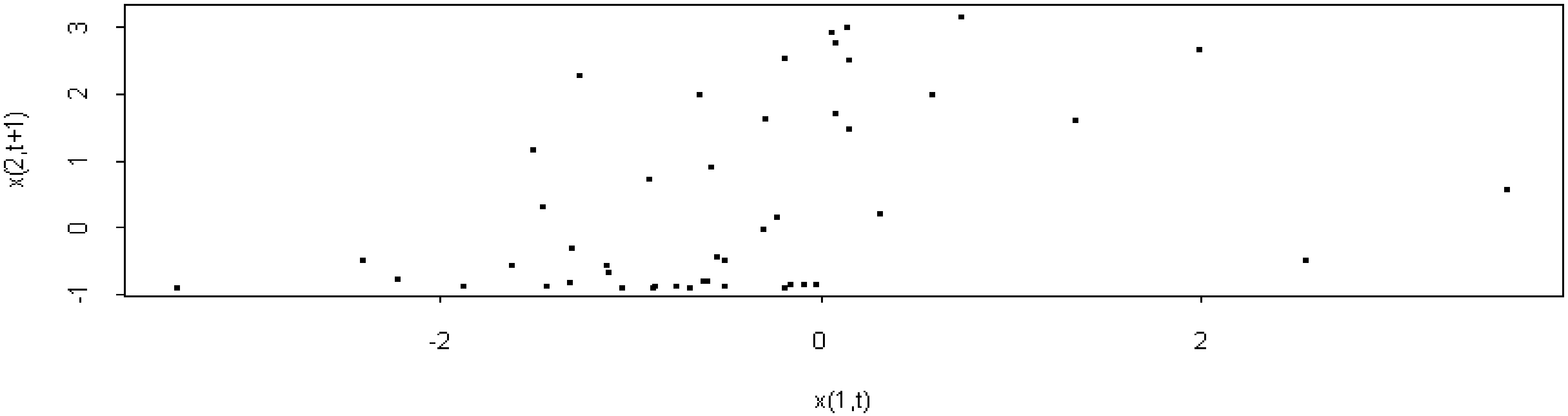
- Simulate the vector autoregressive model to obtain $z_{i,t}$

Step 2

- Using $z_{i,t}$, generate Johnson variates $x_{i,t} = \varepsilon_i + \lambda_i \, g_i^{-1}\left(\dfrac{z_{i,t} - \gamma_i}{\delta_i}\right)$

Example (cont'd)





Summary and Future Research

- Provided a general-purpose tool for modeling and generating dependent and multivariate input processes

- Given historical data
  - determine which Johnson system to use
  - estimate the parameters of the corresponding distribution
  - capture the dependence structure in the multivariate input data