Automatically Derived Full Conditionals for Factored Regression Models with Missing Data and Latent Variables

Brian T. Keller

Assistant Professor
Statistics, Measurement, & Evaluation in Education
Educational, School, & Counseling Psychology
University of Missouri

Developer, Blimp

Educational, School and
Counseling Psychology
University of Missouri

What is Factored Regression Modeling?

- Factored regression models come from the missing data literature, where multivariate distributions are required to model incomplete predictors.
- They can be useful for specifying imputation models for incomplete predictors that are congenial with the analysis model.

Factored Regression Models

Suppose, we are interested in modeling a three variable problem with X, Y, and Z.

Their joint distribution are represented symbolically as:

$$f(X, Y, Z)$$

Often, we may assume that all three variables follow a multivariate normal distribution with some unknown mean vector ($\boldsymbol{\mu}$) and covariance matrix ($\boldsymbol{\Sigma}$):

$$f(X, Y, Z) = \mathcal{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

As an alternative to modeling the joint distribution, we can model this density using conditional models.

$$f(X, Y, Z) = f\left(Y \mid Z, X\right) \times f\left(Z \mid X\right) \times f(X)$$

We then specify models for these three distributions. For example, a saturated covariance matrix can be modeled via three linear regressions:

$$
\begin{aligned}
f(Y \mid X, Z) &\rightarrow \qquad y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + r_{Yi} \\
f(Z \mid X) &\rightarrow \qquad z_i = \alpha_0 + \alpha_1 x_i + r_{Zi} \\
f(X) &\rightarrow \qquad x_i = \gamma_0 + r_{Xi}
\end{aligned}
$$

Derivation of Conditional Distribution for X

Under conditional normality for the two densities in the product

$$f\left(Y \mid X, \ldots\right) \times f\left(X \mid \ldots\right)$$

if we rearrange the two densities into the following forms

$$\begin{aligned} f\left(Y \mid X, \ldots\right) \rightarrow &\qquad y_i \sim \mathcal{N}\left(\mathrm{A}_i + \mathrm{B}_i x_i, \sigma_e^2\right) \\ f\left(X \mid \ldots\right) \rightarrow &\qquad x_i \sim \mathcal{N}\left(\mathrm{C}_i, \sigma_r^2\right) \end{aligned}$$

where A_i, B_i, and C_i are any computation that does not include x_i or y_i, then we can solve for $f\left(X \mid Y, \ldots\right)$.

$$
\begin{aligned}
f\left(X \mid Y, \ldots\right) &\propto \exp\left(-\frac{\left(y_i - [\mathrm{A}_i + \mathrm{B}_i x_i]\right)^2}{2\sigma_e^2}\right) \times \exp\left(-\frac{\left(x_i - \mathrm{C}_i\right)^2}{2\sigma_r^2}\right) \\
&\propto \exp\left(-\left[\frac{\left(\left[y_i - \mathrm{A}_i\right] - \mathrm{B}_i x_i\right)^2}{2\sigma_e^2} + \frac{\left(x_i - \mathrm{C}_i\right)^2}{2\sigma_r^2}\right]\right) \\
&\propto \exp\left(-\left[\frac{x_i^2 \mathrm{B}_i^2 - 2x_i \mathrm{B}_i \left(y_i - \mathrm{A}_i\right) + \text{const}}{2\sigma_e^2} + \frac{x_i^2 - 2x_i \mathrm{C}_i + \text{const}}{2\sigma_r^2}\right]\right) \\
&\propto \exp\left(-\left[\frac{x_i^2 \mathrm{B}_i^2 - 2x_i \mathrm{B}_i \left(y_i - \mathrm{A}_i\right)}{2\sigma_e^2} + \frac{x_i^2 - 2x_i \mathrm{C}_i}{2\sigma_r^2}\right]\right) \\
&\propto \exp\left(-\frac{x_i^2 \mathrm{B}_i^2 \sigma_r^2 - 2x_i \mathrm{B}_i \sigma_r^2 \left(y_i - \mathrm{A}_i\right) + x_i^2 \sigma_e^2 - 2x_i \sigma_e^2 \mathrm{C}_i}{2\sigma_e^2 \sigma_r^2}\right)
\end{aligned}
$$

$$
\begin{aligned}
f\left(X \mid Y, \ldots\right) &\propto \exp\left(-\frac{x_i^2\left[\mathrm{B}_i^2\sigma_r^2+\sigma_e^2\right]-2x_i\left[\mathrm{B}_i\sigma_r^2\left(y_i-\mathrm{A}_i\right)+\sigma_e^2\mathrm{C}_i\right]}{2\sigma_e^2\sigma_r^2}\right)\\
&\propto \exp\left(-\frac{x_i^2-\frac{2x_i\left[\mathrm{B}_i\sigma_r^2\left(y_i-\mathrm{A}_i\right)+\sigma_e^2\mathrm{C}_i\right]}{\mathrm{B}_i^2\sigma_r^2+\sigma_e^2}}{\frac{2\sigma_e^2\sigma_r^2}{\mathrm{B}_i^2\sigma_r^2+\sigma_e^2}}\right)\\
&\propto \exp\left(-\frac{x_i^2-\frac{2x_i\left[\mathrm{B}_i\sigma_r^2\left(y_i-\mathrm{A}_i\right)+\sigma_e^2\mathrm{C}_i\right]}{\mathrm{B}_i^2\sigma_r^2+\sigma_e^2}+\mathrm{const}}{\frac{2\sigma_e^2\sigma_r^2}{\mathrm{B}_i^2\sigma_r^2+\sigma_e^2}}\right)\\
&\propto \exp\left(-\frac{\left[x_i-\frac{\sigma_r^2\mathrm{B}_i\left(y_i-\mathrm{A}_i\right)+\sigma_e^2\mathrm{C}_i}{\sigma_r^2\mathrm{B}_i^2+\sigma_e^2}\right]^2}{2\left(\frac{\sigma_e^2\sigma_r^2}{\mathrm{B}_i^2\sigma_r^2+\sigma_e^2}\right)}\right)
\end{aligned}
$$

With two densities:

$$f\left(Y \mid X, \dots\right) \rightarrow \qquad y_i \sim \mathcal{N}\left(\mathrm{A}_i + \mathrm{B}_i x_i, \sigma_e^2\right)$$
$$f\left(X \mid \dots\right) \rightarrow \qquad x_i \sim \mathcal{N}\left(\mathrm{C}_i, \sigma_r^2\right)$$

The conditional distribution for X is given by

$$f\left(X \mid Y, \dots\right) = \; x_i \sim \mathcal{N}\left(\frac{\sigma_r^2 \mathrm{B}_i \left(y_i - \mathrm{A}_i\right) + \sigma_e^2 \mathrm{C}_i}{\sigma_r^2 \mathrm{B}_i^2 + \sigma_e^2}, \frac{\sigma_e^2 \sigma_r^2}{\sigma_r^2 \mathrm{B}_i^2 + \sigma_e^2}\right)$$

Example 1: Imputation of Incomplete Variables with Products

Suppose we have the following factored regression with X partially observed:

$$
\begin{aligned}
f\left(Y \mid X, Z\right) \rightarrow &\qquad y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 (x_i \times z_i) + r_{Yi} \\
f\left(Z \mid X\right) \rightarrow &\qquad z_i = \alpha_0 + \alpha_1 x_i + r_{Zi} \\
f(X) \rightarrow &\qquad x_i = \gamma_0 + r_{Xi}
\end{aligned}
$$

Example 1: Imputation of Incomplete Variables with Products

Suppose we have the following factored regression with X partially observed:

$$
\begin{aligned}
f(Y \mid X, Z) &\rightarrow \quad y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 (x_i \times z_i) + r_{Yi} \\
f(Z \mid X) &\rightarrow \quad z_i = \alpha_0 + \alpha_1 x_i + r_{Zi} \\
f(X) &\rightarrow \quad x_i = \gamma_0 + r_{Xi}
\end{aligned}
$$

If our goal is to obtain the distribution for $f(X \mid Y, Z)$ —e.g., as required for a Gibbs sampler to impute X—we first break down this product into two separate steps.

- Obtain $f(X \mid Z) \propto f(Z \mid X) \times f(X)$
- Obtain $f(X \mid Y, Z) \propto f(Y \mid X, Z) \times f(X \mid Z)$

Step 1: Obtain $f\left(X \mid Z\right) \propto f\left(Z \mid X\right) \times f(X)$

Based on the two models we have:

$$f\left(Z \mid X\right) \rightarrow \qquad z_i = \alpha_0 + \alpha_1 x_i + r_{Zi}$$

$$f(X) \rightarrow \qquad x_i = \gamma_0 + r_{Xi}$$

By applying the general derivation we obtain:

$$f\left(X \mid Z\right) = \quad x_i \sim \mathcal{N}\left(\frac{\sigma^2_X \alpha_1 \left(z_i - \alpha_0\right) + \sigma^2_Z \gamma_0}{\sigma^2_X \alpha^2_1 + \sigma^2_Z}, \frac{\sigma^2_Z \sigma^2_X}{\sigma^2_X \alpha^2_1 + \sigma^2_Z}\right)$$

Step 2: Obtain $f\left(X \mid Y, Z\right) \propto f\left(Y \mid X, Z\right) \times f\left(X \mid Z\right)$

Based on Step 1 and the remaining model we have:

$$f\left(Y \mid X, Z\right) \rightarrow \qquad y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 (x_i \times z_i) + r_{Yi}$$

$$f\left(X \mid Z\right) \rightarrow \qquad x_i \sim \mathcal{N}\left(\frac{\sigma^2_X \alpha_1 \left(z_i - \alpha_0\right) + \sigma^2_Z \gamma_0}{\sigma^2_X \alpha^2_1 + \sigma^2_Z}, \frac{\sigma^2_Z \sigma^2_X}{\sigma^2_X \alpha^2_1 + \sigma^2_Z} \right)$$

Step 2: Obtain $f(X \mid Y, Z) \propto f(Y \mid X, Z) \times f(X \mid Z)$

Based on Step 1 and the remaining model we have:

$$f(Y \mid X, Z) \rightarrow \qquad y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 (x_i \times z_i) + r_{Yi}$$

$$f(X \mid Z) \rightarrow \qquad x_i \sim \mathcal{N}\left(\frac{\sigma_X^2 \alpha_1 (z_i - \alpha_0) + \sigma_Z^2 \gamma_0}{\sigma_X^2 \alpha_1^2 + \sigma_Z^2}, \frac{\sigma_Z^2 \sigma_X^2}{\sigma_X^2 \alpha_1^2 + \sigma_Z^2}\right)$$

Rearrange $f(Y \mid X, Z)$ to map onto $y_i = \mathrm{A}_i + \mathrm{B}_i x_i + r_i$:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 (x_i \times z_i) + r_{Yi}$$

Rearrange $f(Y \mid X, Z)$ to map onto $y_i = \mathrm{A}_i + \mathrm{B}_i x_i + r_i$:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 (x_i \times z_i) + r_{Yi}$$

$$= \left(\beta_0 + \beta_2 z_i\right) + \left(\beta_1 + \beta_3 z_i\right) x_i + r_{Yi}$$

Rearrange $f(Y \mid X, Z)$ to map onto $y_i = \mathrm{A}_i + \mathrm{B}_i x_i + r_i$:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 (x_i \times z_i) + r_{Yi}$$

$$= \underbrace{(\beta_0 + \beta_2 z_i)}_{\mathrm{A}_i} + \underbrace{(\beta_1 + \beta_3 z_i)}_{\mathrm{B}_i} x_i + r_{Yi}$$

By applying the derivation we can obtain the conditional distribution for X:

$$f\left(X \mid Y, Z\right) \rightarrow \qquad x_i \sim \mathcal{N}\left(\mu_{X|Y,Z}, \sigma^2_{X|Y,Z}\right)$$

where the mean and variances are as follows.

$$\mu_{X|Y,Z} = \frac{\left(\frac{\sigma_Z^2\sigma_X^2}{\sigma_X^2\alpha_1^2+\sigma_Z^2}\right)\left(\beta_1 + \beta_3 z_i\right)\left(y_i - \left[\beta_0 + \beta_2 z_i\right]\right) + \sigma_Y^2\left(\frac{\sigma_X^2\alpha_1(z_i-\alpha_0)+\sigma_Z^2\gamma_0}{\sigma_X^2\alpha_1^2+\sigma_Z^2}\right)}{\left(\frac{\sigma_Z^2\sigma_X^2}{\sigma_X^2\alpha_1^2+\sigma_Z^2}\right)\left(\beta_1 + \beta_3 z_i\right)^2 + \sigma_Y^2}$$

$$\sigma^2_{X|Y,Z} = \frac{\sigma_Y^2\left(\frac{\sigma_Z^2\sigma_X^2}{\sigma_X^2\alpha_1^2+\sigma_Z^2}\right)}{\left(\frac{\sigma_Z^2\sigma_X^2}{\sigma_X^2\alpha_1^2+\sigma_Z^2}\right)\left(\beta_1 + \beta_3 z_i\right)^2 + \sigma_Y^2}$$

```r
compute_dist <- function(y, a, b, mu_x, var_y, var_x) {
    # Obtain substitutions
    y <- substitute(y)
    a <- substitute(a)
    b <- substitute(b)
    mu_x <- substitute(mu_x)
    var_y <- substitute(var_y)
    var_x <- substitute(var_x)
    # Check if they exist and obtain values, otherwise use name
    y <- tryCatch(eval(y), error = \(.) paste('(',deparse(y),')'))
    a <- tryCatch(eval(a), error = \(.) paste('(',deparse(a),')'))
    b <- tryCatch(eval(b), error = \(.) paste('(',deparse(b),')'))
    mu_x <- tryCatch(eval(mu_x), error = \(.) paste('(',deparse(mu_x),')'))
    var_y <- tryCatch(eval(var_y), error = \(.) paste('(',deparse(var_y),')'))
    var_x <- tryCatch(eval(var_x), error = \(.) paste('(',deparse(var_x),')'))
```

⋮

⋮

```
  # Return Solution
  list(
    mean = paste(
      '(', var_x, '*', b, '*', '(', y, '-', a, ')', '+',
      var_y, '*', mu_x, ')', '/',
      '(', var_x, '*', b, '*', b, '+', var_y, ')'
    ),
    variance = paste(
      '(', var_y, '*', var_x, ')', '/', '(',
      var_x, '*', b, '*', b, '+', var_y, ')'
    )
  )
}
```

```r
fx_z <- compute_dist(
    y = z_i,
    a = a_0,
    b = a_1,
    var_y = s_z^2,
    mu_x = g_0,
    var_x =  s_x^2
)
fx_z |> lapply(str2lang)
```

```
$mean
((s_x^2) * (a_1) * ((z_i) - (a_0)) + (s_z^2) * (g_0))/((s_x^2) *
    (a_1) * (a_1) + (s_z^2))


$variance
((s_z^2) * (s_x^2))/((s_x^2) * (a_1) * (a_1) + (s_z^2))
```

```r
fx_yz <- compute_dist(
    y = y_i,
    a = b_0 + b_2 * z_i,
    b = b_1 + b_3 * z_i,
    var_y = s_x^2,
    mu_x = fx_z$mean,
    var_x = fx_z$variance
)
fx_yz |> lapply(str2lang)
```

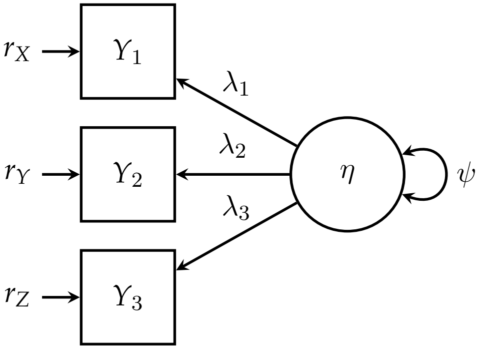
```
$mean
(((s_z^2) * (s_x^2))/((s_x^2) * (a_1) * (a_1) + (s_z^2)) * (b_1 +
    b_3 * z_i) * ((y_i) - (b_0 + b_2 * z_i)) + (s_x^2) * ((s_x^2) *
    (a_1) * ((z_i) - (a_0)) + (s_z^2) * (g_0))/((s_x^2) * (a_1) *
    (a_1) + (s_z^2)))/(((s_z^2) * (s_x^2))/((s_x^2) * (a_1) *
    (a_1) + (s_z^2)) * (b_1 + b_3 * z_i) * (b_1 + b_3 * z_i) +
    (s_x^2))

$variance
((s_x^2) * ((s_z^2) * (s_x^2))/((s_x^2) * (a_1) * (a_1) + (s_z^2)))/(((s_z^2) *
    (s_x^2))/((s_x^2) * (a_1) * (a_1) + (s_z^2)) * (b_1 + b_3 *
    z_i) * (b_1 + b_3 * z_i) + (s_x^2))
```

Example 2: Latent Variable Models

Measurement Models as Factored Regressions

Suppose our trivariate example loads onto a single latent factor:



Factored regression explicitly models the joint distribution of the indicators and latent factor.

$$f(X, Y, Z, \eta) = f\left(X \mid Y, Z, \eta\right) \times f\left(Y \mid Z, \eta\right) \times f\left(Z \mid \eta\right) \times f(\eta)$$

$$f(X \mid \eta) \rightarrow \qquad x_i = \nu_1 + \lambda_1 \eta_i + r_{Xi}$$

$$f(Y \mid \eta) \rightarrow \qquad y_i = \nu_2 + \lambda_2 \eta_i + r_{Yi}$$

$$f(Z \mid \eta) \rightarrow \qquad z_i = \nu_3 + \lambda_3 \eta_i + r_{Zi}$$

$$f(\eta) \rightarrow \qquad \eta_i = \alpha + \zeta_i$$

```r
# Step 1: f(eta | Y_1) propto f(Y_1 | eta) * f(eta)
dist <- compute_dist(
    y = Y_1,
    a = nu_1,
    b = l_1,
    var_y = s_e1^2,
    mu_x = alpha,
    var_x = psi2
)

# Step 2:  f(Y_2 | eta) * dist
dist <- compute_dist(
    y = Y_2,
    a = nu_2,
    b = l_2,
    var_y = s_e2^2,
    mu_x = dist$mean,
    var_x = dist$variance
)

# Step 3: f(Y_3 | eta) * dist
dist <- compute_dist(
    y = Y_3,
    a = nu_3,
    b = l_3,
    var_y = s_e3^2,
    mu_x = dist$mean,
    var_x = dist$variance
)
dist |> lapply(str2lang)
```

```
$mean
(((s_e2^2) * ((s_e1^2) * (psi2))/((psi2) * (l_1) * (l_1) + (s_e1^2)))/(((s_e1^2) *
    (psi2))/((psi2) * (l_1) * (l_1) + (s_e1^2)) * (l_2) * (l_2) +
    (s_e2^2)) * (l_3) * ((Y_3) - (nu_3)) + (s_e3^2) * (((s_e1^2) *
    (psi2))/((psi2) * (l_1) * (l_1) + (s_e1^2)) * (l_2) * ((Y_2) -
    (nu_2)) + (s_e2^2) * ((psi2) * (l_1) * ((Y_1) - (nu_1)) +
    (s_e1^2) * (alpha))/((psi2) * (l_1) * (l_1) + (s_e1^2)))/(((s_e1^2) *
    (psi2))/((psi2) * (l_1) * (l_1) + (s_e1^2)) * (l_2) * (l_2) +
    (s_e2^2)))/(((s_e2^2) * ((s_e1^2) * (psi2))/((psi2) * (l_1) *
    (l_1) + (s_e1^2)))/(((s_e1^2) * (psi2))/((psi2) * (l_1) *
    (l_1) + (s_e1^2)) * (l_2) * (l_2) + (s_e2^2)) * (l_3) * (l_3) +
    (s_e3^2))

$variance
((s_e3^2) * ((s_e2^2) * ((s_e1^2) * (psi2))/((psi2) * (l_1) *
    (l_1) + (s_e1^2)))/(((s_e1^2) * (psi2))/((psi2) * (l_1) *
    (l_1) + (s_e1^2)) * (l_2) * (l_2) + (s_e2^2)))/(((s_e2^2) *
    ((s_e1^2) * (psi2))/((psi2) * (l_1) * (l_1) + (s_e1^2)))/(((s_e1^2) *
    (psi2))/((psi2) * (l_1) * (l_1) + (s_e1^2)) * (l_2) * (l_2) +
    (s_e2^2)) * (l_3) * (l_3) + (s_e3^2))
```

Demonstration of Effective Sample Size Differences

	solved	metropolis	solved/metropolis
`y1.b1`	1116.13	826.57	1.35
`y1.b2`	904.79	911.22	0.99
`y1.s_e`	875.97	642.17	1.36
`y2.b1`	1226.99	927.26	1.32
`y2.b2`	1291.22	899.91	1.44
`y2.s_e`	1233.07	930.87	1.32
`y3.b1`	1195.12	573.24	2.08
`y3.b2`	1041.56	739.51	1.41
`y3.s_e`	870.15	782.96	1.11

Example 3: Incomplete Dynamical SEM with Latent Centering

AR1 Model:

$$y_{(t)i} = \beta_i + \rho_y(y_{(t-1)i} - \beta_i) + e_{(t)i}$$

Factorization:

$$f\Big(Y_{(t)} \mid Y_{(t+1)}, Y_{(t-1)}\Big) \propto f\Big(Y_{(t+1)} \mid Y_{(t)}\Big) \times f\Big(Y_{(t)} \mid Y_{(t-1)}\Big)$$

$$f\left(Y_{(t+1)} \mid Y_{(t)}\right) \rightarrow \qquad y_{(t+1)i} = \beta_i + \rho_y(y_{(t)i} - \beta_i) + e_{(t+1)i}$$

$$f\left(Y_{(t)} \mid Y_{(t-1)}\right) \rightarrow \qquad y_{(t)i} = \beta_i + \rho_y(y_{(t-1)i} - \beta_i) + e_{(t)i}$$

$$f\left(Y_{(t+1)} \mid Y_{(t)}\right) \rightarrow \qquad y_{(t+1)i} = \beta_i + \rho_y(y_{(t)i} - \beta_i) + e_{(t+1)i}$$

$$f\left(Y_{(t)} \mid Y_{(t-1)}\right) \rightarrow \qquad y_{(t)i} = \beta_i + \rho_y(y_{(t-1)i} - \beta_i) + e_{(t)i}$$

We can rearrange to A, B, and C from the derivation:

- $A_{(t)i} = \beta_i - \rho_y \beta_i$
- $B_{(t)i} = \rho_y$
- $C_{(t)i} = \beta_i + \rho_y(y_{(t-1)i} - \beta_i)$

Parallel Process Model:

$$y_{(t)i} = \beta_{0i} + \beta_1(x_{(t-1)i} - \gamma_{0i}) + \rho_y(y_{(t-1)i} - \beta_{0i}) + e_{(t)i}$$

$$x_{(t)i} = \gamma_{0i} + \gamma_1(y_{(t-1)i} - \beta_{0i}) + \rho_x(x_{(t-1)i} - \gamma_{0i}) + r_{(t)i}$$

Factorization:

$$f\left(Y_{(t)} \mid \ldots\right) \propto f\left(Y_{(t+1)} \mid Y_{(t)}, X_{(t)}\right) \times f\left(Y_{(t)} \mid Y_{(t-1)}, X_{(t-1)}\right) \times f\left(X_{(t+1)} \mid Y_{(t)}, X_{(t)}\right)$$

R scripts available at:

`https://github.com/blimp-stats/psychoco-2025`

Blimp Software

- General-purpose Bayesian estimation for regression and path models.
- Single and multilevel models
- Allows for latent variables, incomplete predictors and outcomes
- Interactive and nonlinear effects
- Nonnormal data
- And more!

Freely available at

`https://www.appliedmissingdata.com/blimp`