1 Probability assignment and inferring transactions from financial statements

Suppose a financial statement analyst has identified a platform, A, for the double entry accounting system describing the analyst's perception of an organization's financial reporting such that

$$Ay = x$$

where A is an assigned $m \times n$ matrix of simple journal entries in the columns, one 1 (denoting a debit entry) and one -1 (denoting a credit entry) in each column, with rows identifying the account adjusted, y is an n-element vector of unknown (to the analyst) transactions amounts to be inferred, and x is an m-element vector of observed changes in account balances over the reporting period. Given a platform A and financial statement changes in account balances, x, the analyst knows considerable about y. However, typically there are many consistent solutions for y. A general description of these solutions is

$$y = y^p + N^T k$$

where y^p is any consistent solution, N is an $(n - m + 1) \times n$ matrix describing a basis for the nullspace of A ($AN^T = 0$), and k is an $(n - m + 1)$-element vector of arbitrary weights on the rows of N. The abundance of consistent solutions is reinforced if we return to the original expression and substitute for y

$$\begin{aligned} Ay &= x \\ A\left(y^p + N^T k\right) &= x \\ Ay^p + AN^T k &= x \\ Ay^p + 0k &= x \\ Ay^p &= x \end{aligned}$$

Since the nullspace is orthogonal (unrelated) to the rows of A, the financial statements convey no information about the null component of y, $N^T k$, and the weights can take on any value. How does the analyst characterize her state of knowledge regarding the transactions in which the firm engaged, y, given platform A and changes in account balance x?

One consistent approach involves assigning prior beliefs regarding y based on the analyst's background knowledge and updating via Bayesian revision. This initial step of assigning prior probability beliefs is critical. Background knowledge can vary from nearly uninformed to highly knowledgeable. For instance, perhaps the analyst has only a weak sense of the location (mean, μ) and an upper bound on the variability (variance, σ^2) for the transactions. Then, the analyst's natural (maximum entropy) probability assignment is a multivariate normal distribution with mean vector μ and variance matrix $\sigma^2 I_n$.[1] On the

[1] Exchangeability or independence is a maximum entropy assignment. Hence, without background knowledge to the contrary, independence is the natural probability assignment.

other hand, a knowledgeable analyst incorporates implications into her priors regarding y. Such knowledge reflects short- or long-run equilibrium strategies based on in-depth knowledge of the organization and its relationships with customers, suppliers, employees, etc. If these implications are summarized by first and second moments for y, the analyst again assigns a normal distribution although in this case a more informed distribution.

Since the weakly informed case is more geometric and intuitive, we develop it first and later discuss the general case including strongly informed.

1.1 Special case: weakly background knowledge

Suppose the analyst assigns prior probability beliefs

$$y \sim \mathbb{N}\left(\mu, \sigma^2 I_n\right)$$

Then

$$x_0 \sim \mathbb{N}\left(A_0\mu, \sigma^2 A_0 A_0^T\right)$$

where A_0 is constructed by dropping any redundant row from A (A_0 has linearly independent rows) and x_0 drops the corresponding element from x. Further, the covariance between y and x_0 is

$$\begin{aligned} E\left[(y-\mu)(x_0 - A_0\mu)^T\right] &= E\left[(y-\mu)(A_0 y - A_0\mu)^T\right] \\ &= E\left[(y-\mu)(y-\mu)^T\right] A_0^T \\ &= \sigma^2 A_0^T \end{aligned}$$

and the covariance between x_0 and y is

$$\begin{aligned} E\left[(x_0 - A_0\mu)(y-\mu)^T\right] &= E\left[(A_0 y - A_0\mu)(y-\mu)^T\right] \\ &= A_0 E\left[(y-\mu)(y-\mu)^T\right] \\ &= \sigma^2 A_0 \end{aligned}$$

The conditional expectation of transactions given platform A and $x = x^0$ is

$$\begin{aligned} E\left[y \mid x = x^0, A\right] &= \mu + \sigma^2 A_0^T \left(\sigma^2 A_0 A_0^T\right)^{-1} \left(x^0 - A_0\mu\right) \\ &= \mu + A_0^T \left(A_0 A_0^T\right)^{-1} \left(A_0 y^p - A_0\mu\right) \\ &= \mu + A_0^T \left(A_0 A_0^T\right)^{-1} A_0 \left(y^p - \mu\right) \\ &= P_{R(A)} y^p + \left(I - P_{R(A)}\right)\mu \end{aligned}$$

where $P_{R(A)} = A_0^T \left(A_0 A_0^T\right)^{-1} A_0$, the projection into the rows of A. Iterated expectations provides a consistency check.

$$\begin{aligned} E[y] &= E_x\left[E\left[y \mid x = x^0, A\right]\right] \\ &= E_x\left[P_{R(A)}y^p + \left(I - P_{R(A)}\right)\mu\right] \\ &= P_{R(A)}E[y^p] + \left(I - P_{R(A)}\right)\mu \\ &= P_{R(A)}\mu + \left(I - P_{R(A)}\right)\mu \\ &= \mu \end{aligned}$$

The conditional variance of transactions given platform A and x doesn't depend on the realized value x^0.

$$\begin{aligned} Var[y \mid x, A] &= \sigma^2 I - \sigma^2 A_0^T \left(\sigma^2 A_0 A_0^T\right)^{-1} \sigma^2 A_0 \\ &= \sigma^2 \left(I - P_{R(A)}\right) \end{aligned}$$

Variance decomposition provides a consistency check.

$$\begin{aligned} Var[y] &= Var_x[E[y \mid x, A]] + E_x[Var[y \mid x, A]] \\ &= Var_x\left[P_{R(A)}y^p + \left(I - P_{R(A)}\right)\mu\right] + Var[y \mid x, A] \\ &= E\left[\begin{array}{c} \left(P_{R(A)}y^p + \left(I - P_{R(A)}\right)\mu - \mu\right) \\ \left(P_{R(A)}y^p + \left(I - P_{R(A)}\right)\mu - \mu\right)^T \end{array} \mid x, A\right] \\ &\quad + Var[y \mid x, A] \\ &= P_{R(A)}E\left[(y^p - \mu)(y^p - \mu)^T\right] P_{R(A)} + Var[y \mid x, A] \\ &= \sigma^2 P_{R(A)} + \sigma^2 \left(I - P_{R(A)}\right) \\ &= \sigma^2 \end{aligned}$$

If the analyst wishes to write a proper posterior density function, residual uncertainty conditional on platform A and financial statement results x resides with k, the weights on the nullspace of A.

$$k \sim \mathbb{N}\left(\left(NN^T\right)^{-1} N\mu, \sigma^2 \left(NN^T\right)^{-1}\right)$$

and

$$\left(y \mid x = x^0, A\right) = y^{R(A)} + N^T k$$

where $y^{R(A)} = P_{R(A)}y^p$. Putting this all together leads back to the results above.

$$\begin{aligned} E\left[y \mid x = x^0, A\right] &= y^{R(A)} + N^T E\left[k \mid x = x^0, A\right] \\ &= P_{R(A)}y^p + N^T \left(NN^T\right)^{-1} N\mu \end{aligned}$$

where

$$N^T \left(NN^T\right)^{-1} N = I - P_{R(A)}$$

and

$$
\begin{aligned}
& Var\left[y \mid x, A\right] \\
= \; & Var\left[y^{R(A)} + N^T k \mid x, A\right] \\
= \; & E\left[\begin{array}{c} \left(y^{R(A)} + N^T k - y^{R(A)} - N^T \left(NN^T\right)^{-1} N\mu\right) \\ \left(y^{R(A)} + N^T k - y^{R(A)} - N^T \left(NN^T\right)^{-1} N\mu\right)^T \end{array} \mid x, A\right] \\
= \; & E\left[\begin{array}{c} \left(N^T k - N^T \left(NN^T\right)^{-1} N\mu\right) \\ \left(N^T k - N^T \left(NN^T\right)^{-1} N\mu\right)^T \end{array} \mid x, A\right] \\
= \; & N^T E\left[\left(k - \left(NN^T\right)^{-1} N\mu\right)\left(k - \left(NN^T\right)^{-1} N\mu\right)^T \mid x, A\right] N \\
= \; & N^T Var\left[k \mid x =, A\right] N \\
= \; & \sigma^2 N^T \left(NN^T\right)^{-1} N \\
= \; & \sigma^2 \left(I - P_{R(A)}\right)
\end{aligned}
$$

1.2 More general case: strong background knowledge

Suppose the analyst assigns prior probability beliefs

$$y \sim \mathbb{N}\left(\mu, \Sigma\right)$$

where μ is likely different than that above and Σ is a general variance matrix accounting for different variability across transactions and correlation between transactions. Then

$$x_0 \sim \mathbb{N}\left(A_0\mu, A_0\Sigma A_0^T\right)$$

Further, the covariance between y and x_0 is

$$
\begin{aligned}
E\left[\left(y - \mu\right)\left(x_0 - A_0\mu\right)^T\right] &= E\left[\left(y - \mu\right)\left(A_0 y - A_0\mu\right)^T\right] \\
&= E\left[\left(y - \mu\right)\left(y - \mu\right)^T\right] A_0^T \\
&= \Sigma A_0^T
\end{aligned}
$$

and the covariance between x_0 and y is

$$
\begin{aligned}
E\left[\left(x_0 - A_0\mu\right)\left(y - \mu\right)^T\right] &= E\left[\left(A_0 y - A_0\mu\right)\left(y - \mu\right)^T\right] \\
&= A_0 E\left[\left(y - \mu\right)\left(y - \mu\right)^T\right] \\
&= A_0\Sigma
\end{aligned}
$$

The conditional expectation of transactions given platform A and $x = x^0$ is

$$\begin{aligned} E\left[y \mid x = x^0, A\right] &= \mu + \Sigma A_0^T \left(A_0 \Sigma A_0^T\right)^{-1} \left(x^0 - A_0 \mu\right) \\ &= \mu + \Gamma \Gamma^T A_0^T \left(A_0 \Gamma \Gamma^T A_0^T\right)^{-1} \left(A_0 y^p - A_0 \mu\right) \\ &= \Gamma \Gamma^{-1} \mu + \Gamma \left\{ \Gamma^T A_0^T \left(A_0 \Gamma \Gamma^T A_0^T\right)^{-1} A_0 \Gamma \right\} \Gamma^{-1} \left(y^p - \mu\right) \\ &= \Gamma P_{R(A\Gamma)} \Gamma^{-1} y^p + \Gamma \left(I - P_{R(A\Gamma)}\right) \Gamma^{-1} \mu \end{aligned}$$

where $P_{R(A\Gamma)} = \Gamma^T A_0^T \left(A_0 \Gamma \Gamma^T A_0^T\right)^{-1} A_0 \Gamma$, the projection into the rows of $A\Gamma$ and $\Sigma = \Gamma \Gamma^T$ by Cholesky decomposition. Iterated expectations provides a consistency check.

$$\begin{aligned} E[y] &= E_x\left[E\left[y \mid x = x^0, A\right]\right] \\ &= E_x\left[\Gamma P_{R(A\Gamma)} \Gamma^{-1} y^p + \Gamma \left(I - P_{R(A\Gamma)}\right) \Gamma^{-1} \mu\right] \\ &= \Gamma P_{R(A\Gamma)} \Gamma^{-1} E\left[y^p\right] + \Gamma \left(I - P_{R(A\Gamma)}\right) \Gamma^{-1} \mu \\ &= \Gamma P_{R(A\Gamma)} \Gamma^{-1} \mu + \Gamma \left(1 - P_{R(A\Gamma)}\right) \Gamma^{-1} \mu \\ &= \Gamma \Gamma^{-1} \mu = \mu \end{aligned}$$

The conditional variance of transactions given platform A and x doesn't depend on the realized value x^0.

$$\begin{aligned} Var\left[y \mid x, A\right] &= \Sigma - \Sigma A_0^T \left(A_0 \Sigma A_0^T\right)^{-1} A_0 \Sigma \\ &= \Gamma I \Gamma^T - \Gamma \left\{ \Gamma^T A_0^T \left(A_0 \Gamma \Gamma^T A_0^T\right)^{-1} A_0 \Gamma \right\} \Gamma^T \\ &= \Gamma \left(I - P_{R(A\Gamma)}\right) \Gamma^T \end{aligned}$$

Variance decomposition provides a consistency check.

$$\begin{aligned} Var[y] &= Var_x\left[E\left[y \mid x, A\right]\right] + E_x\left[Var\left[y \mid x, A\right]\right] \\ &= Var_x\left[\Gamma P_{R(A\Gamma)} \Gamma^{-1} y^p + \Gamma \left(I - P_{R(A\Gamma)}\right) \Gamma^{-1} \mu\right] + Var\left[y \mid x, A\right] \\ &= E\left[\begin{array}{c} \left(\begin{array}{c} \Gamma P_{R(A\Gamma)} \Gamma^{-1} y^p \\ +\Gamma \left(I - P_{R(A\Gamma)}\right) \Gamma^{-1} \mu - \mu \end{array} \right) \\ \left(\begin{array}{c} \Gamma P_{R(A\Gamma)} \Gamma^{-1} y^p \\ +\Gamma \left(I - P_{R(A\Gamma)}\right) \Gamma^{-1} \mu - \mu \end{array} \right)^T \end{array} \mid x, A \right] \\ &\quad + Var\left[y \mid x, A\right] \\ &= \Gamma P_{R(A\Gamma)} \Gamma^{-1} E\left[\left(y^p - \mu\right)\left(y^p - \mu\right)^T\right] \left(\Gamma P_{R(A\Gamma)} \Gamma^{-1}\right)^T \\ &\quad + Var\left[y \mid x, A\right] \\ &= \Gamma P_{R(A\Gamma)} \Gamma^{-1} \Gamma \Gamma^T \left(\Gamma^T\right)^{-1} P_{R(A\Gamma)} \Gamma^T + \Gamma \left(I - P_{R(A\Gamma)}\right) \Gamma^T \\ &= \Gamma P_{R(A\Gamma)} \Gamma^T + \Gamma \left(I - P_{R(A\Gamma)}\right) \Gamma^T \\ &= \Gamma \Gamma^T = \Sigma \end{aligned}$$

If the analyst wishes to write a proper posterior density function, residual uncertainty conditional on platform A and financial statement results x resides with k, the weights on the nullspace of A, N. Notice the nullspace of $A\Gamma$ can be expressed $N\left(\Gamma^{-1}\right)^{T}$ so that $A\Gamma\Gamma^{-1}N^{T}=AN^{T}=0$.

$$(k \mid x) \sim \mathbb{N}\left(\left(N\left(\Gamma^{-1}\right)^{T}\Gamma^{-1}N^{T}\right)^{-1}N\left(\Gamma^{-1}\right)^{T}\Gamma^{-1}\mu, \left(N\left(\Gamma^{-1}\right)^{T}\Gamma^{-1}N^{T}\right)^{-1}\right)$$

and

$$\Gamma\left(\Gamma^{-1}\left(y \mid x=x^{0}, A\right)\right)=\Gamma\left(\Gamma^{-1}y^{P}+\Gamma^{-1}N^{T}k\right)$$

where $y^{R(A\Gamma)}=P_{R(A\Gamma)}\Gamma^{-1}y^{p}$. Putting this all together leads back to the results above.

$$\begin{aligned}
& E\left[y \mid x=x^{0}, A\right] \\
= \;& \Gamma y^{R(A\Gamma)}+\Gamma\left(\Gamma^{-1}N^{T}E\left[k \mid x=x^{0}, A\right]\right) \\
= \;& \Gamma P_{R(A\Gamma)}\Gamma^{-1}y^{p}+\Gamma\left\{\Gamma^{-1}N^{T}\left(N\left(\Gamma^{-1}\right)^{T}\Gamma^{-1}N^{T}\right)^{-1}N\left(\Gamma^{-1}\right)^{T}\right\}\Gamma^{-1}\mu \\
= \;& \Gamma\left[P_{R(A\Gamma)}\Gamma^{-1}y^{p}+\left(I-P_{R(A\Gamma)}\right)\Gamma^{-1}\mu\right]
\end{aligned}$$

where

$$\Gamma^{-1}N^{T}\left(N\left(\Gamma^{-1}\right)^{T}\Gamma^{-1}N^{T}\right)^{-1}N\left(\Gamma^{-1}\right)^{T}=I-P_{R(A\Gamma)}$$

and

$$
\begin{aligned}
& Var\left[y \mid x, A\right] \\
= \; & Var\left[y^{R(A\Gamma)} + \Gamma\left(\Gamma^{-1}N^T k\right) \mid x, A\right] \\
= \; & E\left[\begin{array}{c}\left(\begin{array}{c} y^{R(A\Gamma)} + \Gamma\left(\Gamma^{-1}N^T k\right) - y^{R(A\Gamma)} \\ -\Gamma\left\{\Gamma^{-1}N^T\left(N\left(\Gamma^{-1}\right)^T\Gamma^{-1}N^T\right)^{-1}N\left(\Gamma^{-1}\right)^T\right\}\Gamma^{-1}\mu \end{array}\right) \\ \left(\begin{array}{c} y^{R(A\Gamma)} + \Gamma\left(\Gamma^{-1}N^T k\right) - y^{R(A\Gamma)} \\ -\Gamma\left\{\Gamma^{-1}N^T\left(N\left(\Gamma^{-1}\right)^T\Gamma^{-1}N^T\right)^{-1}N\left(\Gamma^{-1}\right)^T\right\}\Gamma^{-1}\mu \end{array}\right)^T \end{array} \mid x, A\right] \\
= \; & E\left[\begin{array}{c}\left(\begin{array}{c} \Gamma\left(\Gamma^{-1}N^T k\right) \\ -\Gamma\left\{\Gamma^{-1}N^T\left(N\left(\Gamma^{-1}\right)^T\Gamma^{-1}N^T\right)^{-1}N\left(\Gamma^{-1}\right)^T\right\}\Gamma^{-1}\mu \end{array}\right) \\ \left(\begin{array}{c} \Gamma\left(\Gamma^{-1}N^T k\right) \\ -\Gamma\left\{\Gamma^{-1}N^T\left(N\left(\Gamma^{-1}\right)^T\Gamma^{-1}N^T\right)^{-1}N\left(\Gamma^{-1}\right)^T\right\}\Gamma^{-1}\mu \end{array}\right)^T \end{array} \mid x, A\right] \\
= \; & \Gamma\left(\Gamma^{-1}N^T\right)E\left[\begin{array}{c}\left(k - \left(N\left(\Gamma^{-1}\right)^T\Gamma^{-1}N^T\right)^{-1}N\left(\Gamma^{-1}\right)^T\Gamma^{-1}\mu\right) \\ \left(k - \left(N\left(\Gamma^{-1}\right)^T\Gamma^{-1}N^T\right)^{-1}N\left(\Gamma^{-1}\right)^T\Gamma^{-1}\mu\right)^T \end{array} \mid x, A\right] \\
& \times\left(N\left(\Gamma^{-1}\right)^T\right)\Gamma^T \\
= \; & \Gamma\left\{\Gamma^{-1}N^T Var\left[k \mid x =, A\right]N\left(\Gamma^{-1}\right)^T\right\}\Gamma^T \\
= \; & \Gamma\left\{\Gamma^{-1}N^T\left(N\left(\Gamma^{-1}\right)^T\Gamma^{-1}N^T\right)^{-1}N\left(\Gamma^{-1}\right)^T\right\}\Gamma^T \\
= \; & \Gamma\left(I - P_{R(A\Gamma)}\right)\Gamma^T
\end{aligned}
$$

1.3 Bayesian details

Now, we embellish the above analysis by returning to Bayesian basics to ensure consistency. As uncertainty is entirely reflected in the distribution for y, $y = y^{\text{row}} + y^{null}$, and $A_0 y = x_0$, we work with $\begin{bmatrix} A_0 \\ N \end{bmatrix} y$.

$$
\begin{aligned}
\begin{bmatrix} A_0 \\ N \end{bmatrix} y &= \begin{bmatrix} A_0 \\ N \end{bmatrix}\left(y^{\text{row}} + N^T k\right) \\
&= \begin{bmatrix} A_0 y^{\text{row}} \\ N N^T k \end{bmatrix} \\
&= \begin{bmatrix} x_0 \\ N N^T k \end{bmatrix}
\end{aligned}
$$

The above is without loss of information as a linear transformation recovers the components of y.

$$\begin{bmatrix} A_0^T \left(A_0 A_0^T\right)^{-1} & 0 \\ 0 & N^T \left(NN^T\right)^{-1} \end{bmatrix} \begin{bmatrix} A_0 \\ N \end{bmatrix} y = \begin{bmatrix} A_0^T \left(A_0 A_0^T\right)^{-1} A_0 y \\ N^T \left(NN^T\right)^{-1} Ny \end{bmatrix} = \begin{bmatrix} y^{\text{row}} \\ y^{null} \end{bmatrix}$$

The joint density is $f(A_0 y, Ny)$ and $f(x_0)$ is written $f(A_0 y)$ so that the posterior distribution is

$$\begin{aligned} f(Ny \mid x_0) &= f(Ny \mid A_0 y) \\ &= \frac{f(A_0 y, Ny)}{f(A_0 y)} \end{aligned}$$

The row component of y, y^{row}, is the projection of any consistent solution, y^p, into the rows of $A_0\Gamma$, $\Gamma\Gamma^T = Var[y] \equiv \Sigma$. In the special case where $\Sigma = \sigma^2 I$, $\Gamma = \sigma I$. Since the scalar can be ignored, y^{row} is the row component of A_0. The null component of y, y^{null}, is the orthogonal component to y^{row}.

1.3.1 Special case, $\Sigma = \sigma^2 I$

Suppose background knowledge is summarized as $y \sim \mathbb{N}\left(\mu, \sigma^2 I\right)$, then

$$\begin{bmatrix} A_0 \\ N \end{bmatrix} y \sim \mathbb{N}\left(\begin{bmatrix} A_0 \\ N \end{bmatrix} \mu, \sigma^2 \begin{bmatrix} A_0 A_0^T & A_0 N^T \\ N A_0^T & NN^T \end{bmatrix} = \sigma^2 \begin{bmatrix} A_0 A_0^T & 0 \\ 0 & NN^T \end{bmatrix} \right)$$

$$A_0 y = x_0 \sim \mathbb{N}\left(A_0 \mu, \sigma^2 A_0 A_0^T\right)$$

and

$$f(Ny \mid x_0) = \frac{f(A_0 y, Ny)}{f(A_0 y)}$$

where

$$\begin{aligned} f(A_0 y, Ny) &= \frac{1}{(2\pi)^{\frac{n}{2}} \left| \sigma^2 \begin{bmatrix} A_0 A_0^T & 0 \\ 0 & NN^T \end{bmatrix} \right|^{\frac{1}{2}}} \\ &\quad \exp\left[\begin{array}{c} -\frac{1}{2\sigma^2} (y-\mu)^T \begin{bmatrix} A_0^T & N^T \end{bmatrix} \begin{bmatrix} A_0 A_0^T & 0 \\ 0 & NN^T \end{bmatrix}^{-1} \\ \begin{bmatrix} A_0 \\ N \end{bmatrix} (y-\mu) \end{array} \right] \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma \left|A_0 A_0^T\right|^{\frac{1}{2}} \sigma \left|NN^T\right|^{\frac{1}{2}}} \\ &\quad \exp\left[-\frac{1}{2\sigma^2} (y-\mu)^T \left(\begin{array}{c} A_0^T \left(A_0 A_0^T\right)^{-1} A_0 \\ +N^T \left(NN^T\right)^{-1} N \end{array} \right) (y-\mu) \right] \end{aligned}$$

and

$$f(A_0 y) = \frac{1}{(2\pi)^{\frac{m-1}{2}} \sigma \left|A_0 A_0^T\right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma^2}(y-\mu)^T A_0^T \left(A_0 A_0^T\right)^{-1} A_0 (y-\mu)\right]$$

Hence,

$$f(Ny \mid A_0 y) = \frac{1}{(2\pi)^{\frac{n-m+1}{2}} \sigma \left|NN^T\right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma^2}(y-\mu)^T N^T \left(NN^T\right)^{-1} N (y-\mu)\right]$$

This indicates

$$Var\left[Ny \mid x_0\right] = \sigma^2 NN^T$$

Is this consistent with the analysis above? Recall

$$Ny = NN^T k$$

and from above

$$Var\left[k \mid x_0\right] = \sigma^2 \left(NN^T\right)^{-1}$$

then

$$\begin{aligned} Var\left[NN^T k \mid x_0\right] &= NN^T Var\left[k \mid x_0\right] NN^T \\ &= NN^T \sigma^2 \left(NN^T\right)^{-1} NN^T \\ &= \sigma^2 NN^T \end{aligned}$$

Also, from above

$$Var\left[y \mid x_0\right] = \sigma^2 N^T \left(NN^T\right)^{-1} N$$

which leads to

$$\begin{aligned} Var\left[Ny \mid x_0\right] &= N Var\left[y \mid x_0\right] N^T \\ &= N\left[\sigma^2 N^T \left(NN^T\right)^{-1} N\right] N^T \\ &= \sigma^2 NN^T \end{aligned}$$

Therefore, the conditional variance is consistent with the foregoing analysis.

This leaves the conditional mean. The exponential term includes a quadratic form involving

$$\begin{aligned} N^T \left(NN^T\right)^{-1} N (y-\mu) &= y^{null} - \mu^{null} \\ &= \left(y - y^{\text{row}}\right) - \mu^{null} \\ &= y - \left(y^{\text{row}} + \mu^{null}\right) \\ &= y - E\left[y \mid x_0 = x_0^0\right] \end{aligned}$$

Hence, the conditional expectation is consistent with the foregoing analysis and the Bayesian demonstration for the special case is complete.

1.3.2 general case: $\Sigma \neq \sigma^2 I$

Suppose background knowledge is summarized as $y \sim \mathbb{N}\left(\mu, \Sigma = \Gamma\Gamma^T\right)$, then

$$\Gamma^{-1}y \sim \mathbb{N}\left(\Gamma^{-1}\mu, \Gamma^{-1}\Sigma\left(\Gamma^T\right)^{-1} = I\right)$$

$$A_0\Gamma\Gamma^{-1}y = x_0$$
$$\Gamma^{-1}y = \Gamma^{-1}y^{\text{row}} + \Gamma^{-1}N^T k$$
$$A_0\Gamma\Gamma^{-1}N^T = A_0 N^T = 0$$

which implies $N\left(\Gamma^T\right)^{-1}$ is a basis for the nullspace of $A_0\Gamma$. We proceed in analogous fashion as above by working with

$$\begin{aligned}
&\begin{bmatrix} A_0\Gamma \\ N\left(\Gamma^T\right)^{-1} \end{bmatrix}\Gamma^{-1}y \\
\sim\ & \mathbb{N}\left(\begin{bmatrix} A_0\Gamma \\ N\left(\Gamma^T\right)^{-1} \end{bmatrix}\Gamma^{-1}\mu, \begin{bmatrix} A_0\Gamma\Gamma^T A_0^T & A_0\Gamma\Gamma^{-1}N^T \\ N\left(\Gamma^T\right)^{-1}\Gamma^T A_0^T & N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T \end{bmatrix}\right) \\
=\ & \mathbb{N}\left(\begin{bmatrix} A_0 \\ N\left(\Gamma^T\right)^{-1}\Gamma^{-1} \end{bmatrix}\mu, \begin{bmatrix} A_0\Gamma\Gamma^T A_0^T & 0 \\ 0 & N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T \end{bmatrix}\right)
\end{aligned}$$

$$A_0\Gamma\Gamma^{-1}y = x_0 \sim \mathbb{N}\left(A_0\mu, A_0\Gamma\Gamma^T A_0^T\right)$$

and

$$f\left(N\left(\Gamma^T\right)^{-1}\Gamma^{-1}y \mid x_0\right) = \frac{f\left(A_0\Gamma\Gamma^{-1}y, N\left(\Gamma^T\right)^{-1}\Gamma^{-1}y\right)}{f\left(A_0\Gamma\Gamma^{-1}y\right)}$$

where

$$\begin{aligned}
& f\left(A_0\Gamma\Gamma^{-1}y, N\left(\Gamma^T\right)^{-1}\Gamma^{-1}y\right) \\
=\ & \frac{1}{(2\pi)^{\frac{n}{2}}\left|\begin{bmatrix} A_0\Gamma\Gamma^T A_0^T & 0 \\ 0 & N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T \end{bmatrix}\right|^{\frac{1}{2}}} \\
& \exp\begin{bmatrix} -\frac{1}{2}(y-\mu)^T\left(\Gamma^T\right)^{-1}\begin{bmatrix} \Gamma^T A_0^T & \Gamma^{-1}N^T \end{bmatrix} \\ \begin{bmatrix} A_0\Gamma\Gamma^T A_0^T & 0 \\ 0 & N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T \end{bmatrix}^{-1} \\ \begin{bmatrix} A_0\Gamma \\ N\left(\Gamma^T\right)^{-1} \end{bmatrix}\Gamma^{-1}(y-\mu) \end{bmatrix} \\
=\ & \frac{1}{(2\pi)^{\frac{n}{2}}\left|A_0\Gamma\Gamma^T A_0^T\right|^{\frac{1}{2}}\left|N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T\right|^{\frac{1}{2}}} \\
& \exp\begin{bmatrix} -\frac{1}{2}(y-\mu)^T\left(\Gamma^T\right)^{-1} \\ \left(\begin{array}{c} \Gamma^T A_0^T\left(A_0\Gamma\Gamma^T A_0^T\right)^{-1}A_0\Gamma \\ +\Gamma^{-1}N^T\left(N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T\right)^{-1}N\left(\Gamma^T\right)^{-1} \end{array}\right) \\ \Gamma^{-1}(y-\mu) \end{bmatrix}
\end{aligned}$$

and

$$f\left(A_0\Gamma\Gamma^{-1}y\right) = \frac{1}{(2\pi)^{\frac{m-1}{2}}\left|A_0\Gamma\Gamma^T A_0^T\right|^{\frac{1}{2}}} \exp\left[\begin{array}{c} -\frac{1}{2}(y-\mu)^T\left(\Gamma^T\right)^{-1} \\ \Gamma^T A_0^T\left(A_0\Gamma\Gamma^T A_0^T\right)^{-1} A_0\Gamma \\ \Gamma^{-1}(y-\mu) \end{array}\right]$$

Hence,

$$\begin{aligned} & f\left(N\left(\Gamma^T\right)^{-1}\Gamma^{-1}y \mid A_0\Gamma\Gamma^{-1}y\right) \\ = & \frac{1}{(2\pi)^{\frac{n-m+1}{2}}\left|N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T\right|^{\frac{1}{2}}} \\ & \exp\left[\begin{array}{c} -\frac{1}{2}(y-\mu)^T\left(\Gamma^T\right)^{-1} \\ \Gamma^{-1}N^T\left(N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T\right)^{-1} N\left(\Gamma^T\right)^{-1} \\ \Gamma^{-1}(y-\mu) \end{array}\right] \end{aligned}$$

This indicates

$$Var\left[N\left(\Gamma^T\right)^{-1}\Gamma^{-1}y \mid x_0\right] = N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T$$

Is this consistent with the analysis above? Recall, for the general case

$$N\left(\Gamma^T\right)^{-1}\Gamma^{-1}y = N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T k$$

and from above

$$Var\left[k \mid x_0\right] = \left(N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T\right)^{-1}$$

then

$$\begin{aligned} & Var\left[N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T k \mid x_0\right] \\ = & N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T Var\left[k \mid x_0\right] N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T \\ = & N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T\left(N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T\right)^{-1} \\ & N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T \\ = & N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T \end{aligned}$$

Also, from the general case above

$$Var\left[y \mid x_0\right] = \Gamma\left\{\Gamma^{-1}N^T\left(N\left(\Gamma^{-1}\right)^T\Gamma^{-1}N^T\right)^{-1} N\left(\Gamma^{-1}\right)^T\right\}\Gamma^T$$

which leads to

$$
\begin{aligned}
Var\left[N\left(\Gamma^T\right)^{-1}\Gamma^{-1}y \mid x_0\right] &= N\left(\Gamma^T\right)^{-1}\Gamma^{-1}Var\left[y \mid x_0\right]\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T \\
&= N\left(\Gamma^T\right)^{-1}\Gamma^{-1}\Gamma \\
&\quad \left\{\Gamma^{-1}N^T\left(N\left(\Gamma^{-1}\right)^T\Gamma^{-1}N^T\right)^{-1}N\left(\Gamma^{-1}\right)^T\right\} \\
&\quad \Gamma^T\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T \\
&= N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T
\end{aligned}
$$

Therefore, the conditional variance is consistent with the foregoing analysis.

This leaves the conditional mean. The exponential term includes a quadratic form involving

$$
\begin{aligned}
&\left\{\Gamma^{-1}N^T\left(N\left(\Gamma^T\right)^{-1}\Gamma^{-1}N^T\right)^{-1}N\left(\Gamma^T\right)^{-1}\right\}\Gamma^{-1}\left(y-\mu\right) \\
= \;&\Gamma^{-1}\left(y^{null}-\mu^{null}\right) \\
= \;&\Gamma^{-1}y-\left[P_{R(A\Gamma)}\Gamma^{-1}y+\left(I-P_{R(A\Gamma)}\right)\Gamma^{-1}\mu\right] \\
= \;&\Gamma^{-1}y-E\left[\Gamma^{-1}y \mid x_0=x_0^0\right]
\end{aligned}
$$

where y^{null}, μ^{null} are nullspace (orthogonal) components of $A_0\Gamma$. Hence, the conditional expectation is consistent with the foregoing analysis and the Bayesian demonstration for the general case is complete.