We assume the ARMA(1,1) structure

$$
\begin{aligned}
Cov\left(Y_{ij}, Y_{ik}\right) =& \begin{cases} \sigma_s^2 + \sigma_e^2, & j = k \\ \sigma_s^2 \rho^{|t_{ij} - t_{ik}|}, & j \neq k \end{cases} \\
=& \sigma_s^2 \rho^{|t_{ij} - t_{ik}|} + \sigma_e^2 \delta_{jk}
\end{aligned}
$$

with the time interval in the exponent standardized to $[0, 1]$, and $\delta_{jk} = 1\{j = k\}$. σ_s^2 can be thought of as the variance of the subject-specific random process, and σ_e^2 as the variance of the "nugget" or pure measurement error.

Consider for now only the pairs of observations with $j \neq k$. Let P be the cross-product of residuals from these pairs, and let G be the lag $|t_{ij} - t_{ik}|$, with time standardized. Then

$$
\begin{aligned}
E(P) =& \sigma_s^2 \rho^G \\
\log E(P) =& \log\left(\sigma^2\right) + \log\left(\rho\right) G
\end{aligned}
$$

We can't just estimate $\log\left(\rho\right)$and $\log\left(\sigma_s^2\right)$ from a linear regression of $\log P$ on G. This is both because of Jensen's inequality, and because some of the observed P's will be negative and will not have logarithms. However, we can use the linear relationship to get crude method of moments estimates to use as starting values for nonlinear estimation.

$$
\log \bar{P} = \log\left(\sigma_s^2\right) + \log\left(\rho\right) \bar{G}
$$

We still have a problem: we are trying to solve one equation with two unknowns. So if we want an estimate of $\log\left(\rho\right)$, we first need an estimate of $\log\left(\sigma_s^2\right)$. From the mean cross-product of residuals for the observations with $j = k$, we can easily get an estimate of $\sigma_{total}^2 = \sigma_s^2 + \sigma_e^2$. We do not know how to get σ_s^2 from σ_{total}^2. Just to get a starting value, however, we could assume $\sigma_s^2 = \sigma_e^2$. Then $\tilde{\sigma}_s^2 = \frac{1}{2}\hat{\sigma}_{total}^2$. So after some algebra,

$$
\tilde{\rho} = \exp\left(\frac{\log\left(\bar{P}\right) - \log\left(\tilde{\sigma}_s^2\right)}{\bar{G}}\right).
$$

Unless the model fits very poorly (to the degree that the empirical covariance matrix of the residuals probably would not even be invertible), $\log\left(\bar{P}\right) - \log\left(\tilde{\sigma}_s^2\right) < 0$ so we have $0 < \tilde{\rho} < 1$ as required for interpretability.

Once we have the starting values, we can use Fisher scaling to get more accurate estimates.