# ONLINE RESOURCE 1

for

"Probabilistic Projections of the Total Fertility Rate for All Countries"

Leontine Alkema, Adrian E. Raftery, Patrick Gerland,
Samuel J. Clark, François Pelletier, Thomas Buettner, Gerhard K. Heilig

Demography, August 2011

## Expected decrements in Phase II

During Phase II, for countries that are going through the fertility transition from high fertility toward replacement fertility, we model the changes in the TFR as a function of its level, based on current UN methodology (United Nations, Department of Economic and Social Affairs, Population Division 2006). Specifically, the expected five-year decrements in TFR during the fertility transition are given by the sum of two logistic functions that are evaluated at the TFR level at the start of the period. A logistic function exhibits an S-shape growing from an initial level to an upper or lower asymptote. A logistic function $L(\cdot)$ evaluated at TFR level $f$ can be written as $L(f) = \frac{d_c}{1+\exp\left(-\frac{2\ln(p)}{\triangle}(f-f_{50\%})\right)}$. In this parametrization, the logistic function decreases from upper bound $d_c$ to lower bound 0 as the TFR decreases from high toward low TFR. The midpoint of the decrease is given by TFR level $f_{50\%}$: $L(f_{50\%}) = 0.5d_c$. Parameter $\triangle$ represents the length of the interval in which $L(\cdot)$ decreases from $\frac{p}{p+1}d_c$ to $\frac{1}{p+1}d_c$, thus setting $p = 9$ gives $\triangle = f_{90\%} - f_{10\%}$, which we will refer to as the 80% range of the logistic function.

To model the expected five-year decrements as a function of the TFR, the logistic function above is combined with a second logistic function which describes the opposite force; a logistic function that increases from negative decrement $-d_c$ toward 0 as the TFR decreases (Meyer 1994). The 50% midpoint of the second function is at a higher TFR level than the first function, such that the sum of the two functions is 0 at very high TFR levels, increases toward the maximum decrement $d_c$ as the second logistic increases toward 0, and then decreases toward 0 as the first logistic decreases. The parametrization of the sum of the two logistic functions is as follows (using $p$=9):

$$g(\boldsymbol{\theta}_c, f_{c,t}) = \frac{d_c}{1+\exp\left(-\frac{2\ln(9)}{\triangle_{c3}}(f_{c,t}-\triangle_{c4}-0.5\triangle_{c3})\right)} + \frac{-d_c}{1+\exp\left(-\frac{2\ln(9)}{\triangle_{c1}}(f_{c,t}-\sum_i \triangle_{ci}+0.5\triangle_{c1})\right)},$$

with parameter vector $\boldsymbol{\theta}_c = (\triangle_{c1}, \triangle_{c2}, \triangle_{c3}, \triangle_{c4}, d_c)$, and $\triangle_{ci} \geq 0$ for $i = 1, 2, 3, 4$. It follows that $0.5\triangle_{c3} + \triangle_{c4}$ is the midpoint of the first logistic function, $\triangle_{c3}$ its 80% range, $0.5\triangle_{c1} + \sum_{i=2}^{4} \triangle_{ci}$ the midpoint of the second logistic function, and $\triangle_{c1}$ its 80% range. At high TFR levels, the two logistic functions cancel out each other, thus $g(\boldsymbol{\theta}_c, f_{c,t}) \approx 0$. At TFR level $U_c = \sum_{i=1}^{4} \triangle_{ci}$, the outcome of the second logistic function is $-0.9d_c$ as it has started to increase toward 0. The outcome of the first logistic function is still above $0.9d_c$, thus the decrement $g(\boldsymbol{\theta}_c, f_{c,t})$ at $U_c$ is between 0 and $0.1d_c$. With similar reasoning, it follows that between TFR levels $U_c$ and $U_c - \triangle_{c1}$ the outcome of $g(\boldsymbol{\theta}_c, \cdot)$ increases from around $0.1d_c$ to over $0.8d_c$. During the TFR range $\triangle_{c2}$, the five-year decrements range between $0.8d_c$ and $d_c$, and during $\triangle_{c3}$ the pace of the fertility decline decreases further to $0.1d_c$ at TFR level $\triangle_{c4}$.

## Complete fertility transition model

The complete fertility transition model during Phase II is given by:

$$
\begin{aligned}
f_{c,t+1} &= f_{c,t} - d_{c,t} + \varepsilon_{c,t}, \text{ for } c = 1, \ldots, C, \quad t = \tau_c, \ldots, \lambda_c - 1, \\
\varepsilon_{c,t} &\sim \begin{cases} N(m_t, s_t^2), & \text{for } t = \tau_c, \\ N(0, \sigma(f_{c,t})^2), & \text{otherwise}, \end{cases} \\
d_{c,t} &= \begin{cases} g(\theta_c, f_{c,t}), & \text{for } f_{c,t} > 1, \\ 0, & \text{otherwise}, \end{cases} \\
\theta_c &= (\triangle_{c1}, \triangle_{c2}, \triangle_{c3}, \triangle_{c4}, d_c), \\
g(\theta_c, f_{c,t}) &= \frac{-d_c}{1+\exp\left(-\frac{2\ln(9)}{\triangle_{c1}}(f_{c,t} - \sum_i \triangle_{ci} + 0.5\triangle_{c1})\right)} + \frac{d_c}{1+\exp\left(-\frac{2\ln(9)}{\triangle_{c3}}(f_{c,t} - \triangle_{c4} - 0.5\triangle_{c3})\right)}, \\
U_c & \begin{cases} = f_{c,\tau}, & \text{if } \tau_c \geq 1950-1955; \\ \sim U(\min\{5.5, \max_t f_{c,t}\}, 8.8), & \text{for } \tau_c < 1950-1955. \end{cases}
\end{aligned}
$$

The upper bound of the prior distribution for $U_c$ for countries in which the decline had possibly already started before 1950–1955 is based on the observed maximum in the UN estimates–namely 8.7. Its lower bound is the minimum of the maximum observed TFR value and 5.5 children (5.5 children is based on examining decline curves, the minimum level at which the decline starts is slightly under 6).

The expression for the standard deviation $\sigma(f_{c,t})$ of the distortion terms after start period $\tau_c$ is:

$$
\sigma(f_{c,t}) = c_{1975}(t) \left(\sigma_0 + (f_{c,t} - S)\left(-aI_{[S,\infty)}(f_{c,t}) + bI_{[0,S)}(f_{c,t})\right)\right),
$$

where $\sigma_0$ is the maximum standard deviation of the distortions, attained at TFR level $S$, and $a$ and $b$ are multipliers of the standard deviation, to model the linear decrease for larger and smaller outcomes of the TFR. The constant $c_{1975}(t)$ is added to model the higher error variance of the distortions before 1975, and is given by:

$$
c_{1975}(t) = \begin{cases} c_{1975}, & t \in [1950-1955, 1970-1975]; \\ 1, & t \in [1975-1980, \infty). \end{cases} \tag{1}
$$

The hierarchical part of the transition model is given by:

$$
\begin{aligned}
d_c^* &= \log\left(\frac{d_c - 0.25}{2.5 - d_c}\right), \\
d_c^* &\sim N(\chi, \psi^2), \\
\triangle_{c4}^* &= \log\left(\frac{\triangle_{c4} - 1}{2.5 - \triangle_{c4}}\right), \\
\triangle_{c4}^* &\sim N(\triangle_4, \delta_4^2), \\
p_{ci} &= \frac{\triangle_{ci}}{U_c - \triangle_{c4}} \text{ for } i = 1, 2, 3, \\
p_{ci} &= \frac{\exp(\gamma_{ci})}{\sum_{j=1}^{3} \exp(\gamma_{cj})}, \\
\gamma_{ci} &\sim N(\alpha_i, \delta_i^2).
\end{aligned}
$$

The country-specific parameters in the model are given by $\{\gamma_{ci}, U_c, d_c, \triangle_{c4}\}$, $i = 1, 2, 3$. The hyperparameters in the model are given by $\{\chi, \psi^2, \triangle_4, \delta_4, \boldsymbol{\alpha}, \boldsymbol{\delta}\}$ and $\{a, b, S, \sigma_0, c_{1975}, m_\tau, s_\tau\}$. The prior distributions on the hyperparameters are given by:

$$\begin{aligned}
\chi &\sim N(-1.5, 0.6^2), \\
1/\psi^2 &\sim \text{Gamma}(1, 0.6^2), \\
\alpha_1 &\sim N(-1, 1), \\
\alpha_2 &\sim N(0.5, 1), \\
\alpha_3 &\sim N(1.5, 1), \\
1/\delta_i^2 &\sim \text{Gamma}(1, 1), \text{ for } i = 1, \ldots, 3 \\
1/\delta_4^2 &\sim \text{Gamma}(1, 0.8^2), \\
\triangle_4 &\sim N(0.3, 0.8^2), \\
a &\sim U[0, 0.2], \\
b &\sim U[0, 0.2], \\
\sigma_0 &\sim U[0.01, 0.6], \\
c_{1975} &\sim U[0.8, 2], \\
S &\sim U[3.5, 6.5], \\
m_\tau &\sim N(-0.25, 0.4^2), \\
1/s_\tau^2 &\sim \text{Gamma}(1, 0.4^2).
\end{aligned}$$

The prior distributions on the hyperparameters are chosen based on (i) initial least-squares fits to fertility declines in countries that had observed most of the transition and/or (ii) guesses of reasonable outcomes.

Convergence of all model parameters was assessed using the run length diagnostic of Raftery and Lewis (1992, 1996). The length of the MCMC chain exceeded the required sample size for estimating the 2.5% and 97.5% percentiles of the posterior distributions of all model parameters to within +/-0.0125 accuracy with probability 0.95. Convergence of $\alpha_i$, $i = 1, 2, 3$, was assessed on the transformed scale, i.e. $\alpha_i/(\sum_{j=1}^{3} \alpha_j)$, as these parameters are only weakly identified on their original scale (the likelihood of the data conditional on these parameters, and thus the projections, are not altered when adding a constant to all three $\alpha_i$'s). Similarly for the $\gamma_{ci}$'s, $i = 1, 2, 3$, $c = 1, \ldots, C$, convergence was assessed for $\gamma_{ci}/(\sum_{j=1}^{3} \gamma_{cj})$, $i = 1, 2, 3$, $c = 1, \ldots, C$.

The histograms of the samples from the posterior distributions of the hyperparameters, as well as the prior density functions are given in Figures 1–2. All priors are more spread out than their posterior distributions (i.e. the posterior is determined mostly by the data).

**Fig. 1.** Histograms of the posterior samples of the hierarchical parameters, with prior density function (black).

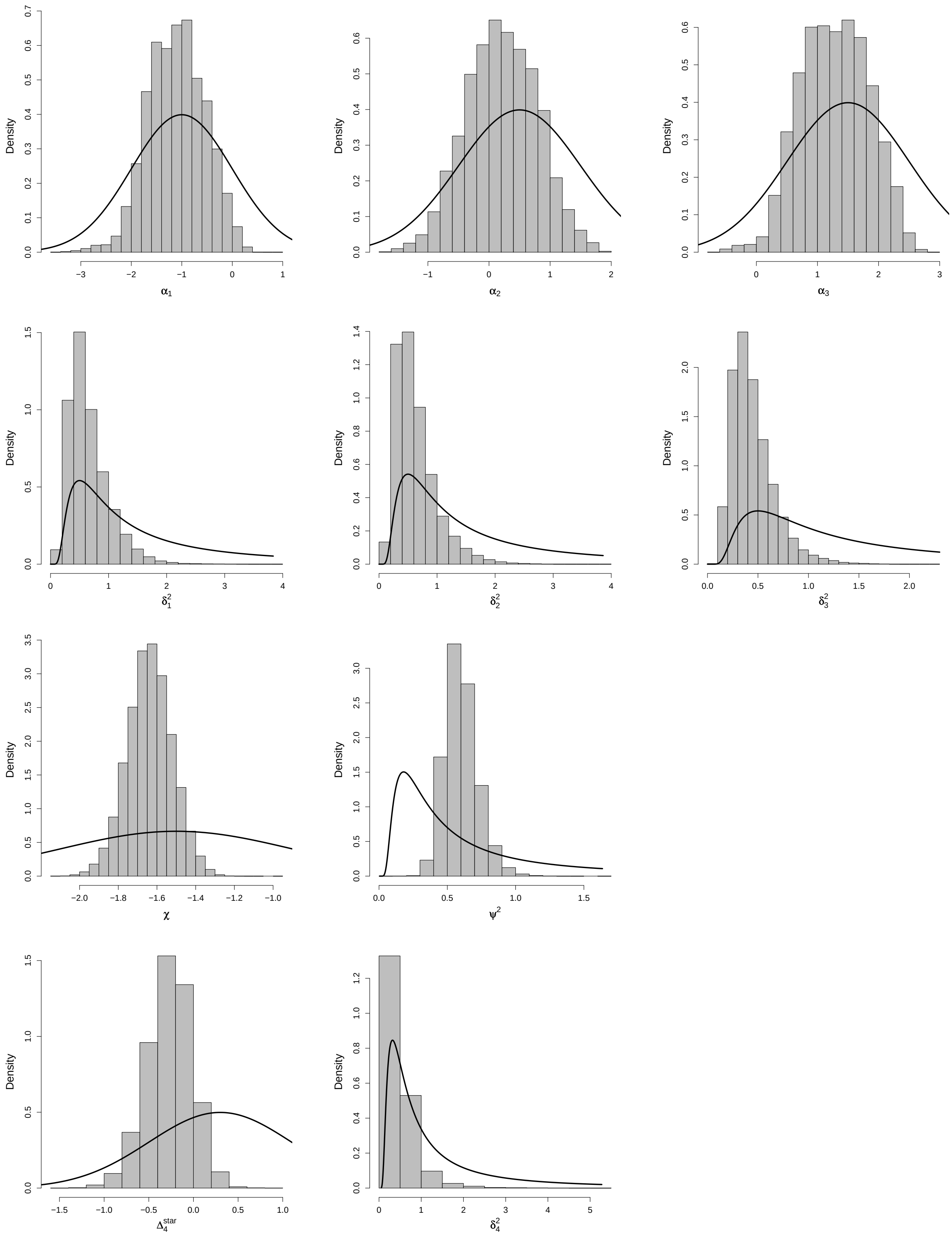
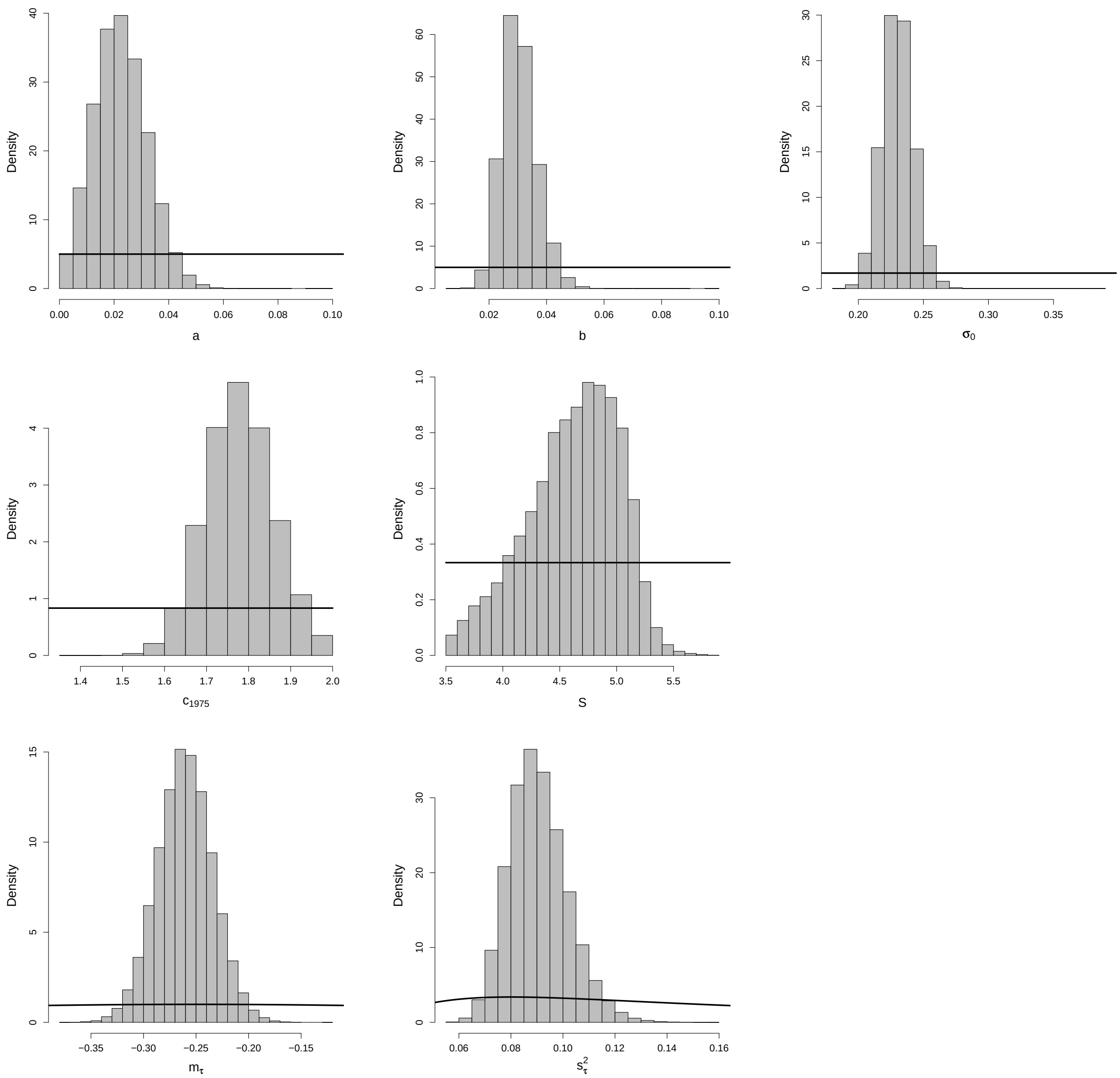


**Fig. 2.** Histograms of the posterior samples of variance parameters of the distortion terms, with prior density function (black).



# References


Meyer, P. (1994). Bi-Logistic Growth. *Technological Forecasting and Social Change 47*, 89–102.

Raftery, A. E. and S. M. Lewis (1992). How many iterations in the Gibbs sampler? In J. M. Bernardo et al. (Ed.), *Bayesian Statistics 4*, pp. 763–773. Oxford University Press.

Raftery, A. E. and S. M. Lewis (1996). Implementing MCMC. In W. R. Gilks, D. J. Spiegelhalter, and S. Richardson (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 115–130. London: Chapman and Hall.

United Nations, Department of Economic and Social Affairs, Population Division (2006). *World Population Prospects. The 2004 Revision, Vol. III, Chapter VI. Methodology of the United Nations population estimates and projections*, pp. 100–104. http://www.un.org/esa/population/publications/WPP2004/WPP2004_Volume3.htm.