Appendixes
Appendix A

Notation

A.1 Principles

Variables and Parameters We use Greek symbols for unknown quantities, such as regression coefficients ($\beta$), expected values ($\mu$), disturbances ($\epsilon$), and variances ($\sigma^2$), and Roman symbols for observed quantities, such as $y$ and $m$ for the dependent variable, while the symbols $X$ and $Z$ refer to covariates.

Parameters that are unknown, but are treated as known rather than estimated, appear in the following font: $\text{abcdef}$. Examples of these user-chosen parameters include the number of derivatives in a smoothing prior ($n$) and some hyperprior parameters (e.g., $f$, $g$).

Indices The indices $i, j = 1, \ldots, N$ refer to generic cross sections. When the cross sections are countries, they may be labeled by the index $c = 1, \ldots, C$; when they are age groups, or specific ages, they may be labeled by the index $a = 1, \ldots, A$. Each cross section also varies over time, which is indexed as $t = 1, \ldots, T$. Cross-sectional time-series variables have the cross-sectional index (or indices) first and the time index last. For example, $m_{it}$ denotes the value of the variable $m$ in cross section $i$ at time $t$, and similarly $m_{cat}$ is the value of the variable $m$ in country $c$ and age group $a$ at time $t$.

Cross section $i$ contains $k_i$ covariates. Therefore $Z_{it}$ is a $k_i \times 1$ vector of covariates and $\beta_i$ is a $k_i \times 1$ vector of coefficients. Every vector or matrix with one or more dimensions equal to $k_i$, such as $Z_{it}$ or $\beta_i$, will be in bold.

Dropping one index from a quantity with one or more indices implies taking the union over the dropped indices, possibly arranging the result in vector form. For example, if $m_{it}$ is the observed value of the dependent variable in cross section $i$ at time $t$, then $m_t$ is an $N \times 1$ column vector whose $j$-th element is $m_{jt}$. We refer to the vector $m_t$ as the cross-sectional profile at time $t$. If the cross sections $i$ are age groups, we call the vector $m_i$ the age profile at time $t$. Applying the same in reverse, we denote by $m_i$ the $T \times 1$ column vector of the time series corresponding to cross section $i$. Iterating this rule results in denoting by $m$ the totality of elements $m_{it}$, and by $\beta$ the totality of vectors $\beta_i$. Similarly, $Z_i$ denotes the standard $T \times k_i$ data matrix for cross section $i$, with rows equal to the vector $Z_{it}$.

If $X$ is a vector, then $\text{diag}(X)$ is the diagonal matrix with $X$ on its diagonal. If $W$ is a matrix, then $\text{diag}(W)$ is the column vector whose elements are the diagonal elements of $W$.

Sums We use the following shorthand for summation whenever it does not create confusion:

$$\sum_{t} \equiv \sum_{t=1}^{T}, \quad \sum_{i} \equiv \sum_{i=1}^{N}, \quad \sum_{c} \equiv \sum_{c=1}^{C}, \quad \sum_{a} \equiv \sum_{a=1}^{A}.$$  

We also define the “summer” vector $\mathbf{1} \equiv (1, 1, \ldots, 1)$ so that for matrix $X$, $X\mathbf{1}$ denotes the row sums.
APPENDIX A

Norms  For a matrix $\mathbf{x}$, we define the weighted Euclidean (or Mahalanobis) norm as $\|\mathbf{x}\|_\Phi^2 \equiv \mathbf{x}^T \Phi \mathbf{x}$, with the standard Euclidean norm as a special case, so that $\|\mathbf{x}\|_I = \|\mathbf{x}\|$, with $I$ as the identity matrix.

Functions  We denote probability densities by capitalized symbols in calligraphic font. For example, the normal density with mean $\mu$ and standard deviation $\sigma$ is $\mathcal{N}(\mu, \sigma^2)$. We denote generic probability densities by $\mathcal{P}$, and for ease of notation we distinguish one density from another only by their arguments. Therefore, for example, instead of writing $\mathcal{P}(\mathbf{x})$ and $\mathcal{P}(\mathbf{z})$, we simply write $\mathcal{P}(\mathbf{x})$ and $\mathcal{P}(\mathbf{z})$.

Sets  Sets such as the real line $\mathbb{R}$ and its subsets ($S \subset \mathbb{R}$) or the natural numbers $\mathbb{N}$ and the integers $\mathbb{Z}$ are denoted with these capital blackboard fonts. We denote the null space of a matrix, operator, or functional as $\mathcal{N}$.

A.2 Glossary

$a$  index for age groups  
$A$  number of age groups 
$b_{it}$  an exogenous weight for an observation at time $t$ in cross section $i$  
$\boldsymbol{\beta}_i$  vector of regression coefficients for cross section $i$  
$\boldsymbol{\beta}_{WLS}^i \equiv (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{y}_i$  the vector of weighted least-squares estimates  
$c$  index for country 
$C$  number of countries 
$d_{it}$  the number of deaths in cross-sectional unit $i$ occurring during time period $t$  
$\delta_{ij}$  Kronecker’s delta function, equal to 1 if $i = j$ and 0 otherwise  
$\mathbb{E}[\cdot]$  the expected value operator  
$\epsilon$  an error term  
$F(\mu)$  summary measures  
$\eta$  an error term  
$i$  index for a generic cross section (with examples being $a$ for age, or $c$ for country)  
$I$  the identity matrix (generic)  
$I_d, I_{d\times d}$  the $d \times d$ identity matrix  
$j$  index for a generic cross section  
$k_i$  the number of covariates in cross section $i$, and the dimension of all corresponding boldface quantities, such as $\boldsymbol{\beta}_i$ and $\mathbf{Z}_{it}$  
$L$  generic diagonal matrix  
$\lambda$  mean of a Poisson event count (section 3.1.1)  
$\ln(\cdot)$  the natural logarithm  
$M_{it}$  mortality rate for cross-sectional unit $i$ at time $t$: $M_{it} \equiv d_{it} / p_{it}$  
$m_{it}$  a generic symbol for the observed value of the dependent variable in cross section $i$ at time $t$. When referring to an application, we use $m_{it} = \ln(M_{it})$, the natural log of the mortality rate.  
$m_{it}$  mean log-mortality age profile, averaging over time, $\overline{m}_a = \frac{\sum_{t=1}^T m_{at}}{T}$
Appendixa.tex

NOTATION

$\tilde{m}$ matrix of mean-centered logged mortality rates, with elements 
$\tilde{m}_{at} \equiv m_{at} - \bar{1}_T \sum_t m_{at}$

$\mu_{it}$ expected value of the dependent variable in cross section $i$ at time $t$

$N$ number of cross-sectional units

$\mathbb{N}$ the set of natural numbers

$n$ generic order of the derivative of the smoothness functional

$\mathbb{R}$ the null space of an operator or a functional

$\mathbb{R}_\perp$ the orthogonal complement of the null space $\mathbb{R}$

$v$ an error term

$O_{q \times d}$ a $q \times d$ matrix of zeros

$p_{it}$ population (number of people) in cross-sectional unit $i$ at the start of time period $t$

$\mathcal{P}$ probability densities. The same $\mathcal{P}$ may refer to two different densities, with the meaning clarified from their arguments.

$Q$ generic correlation matrix of the data

$\mathbb{R}$ the set of real numbers

$s_{ij}$ the weight describing how similar cross-sectional unit $i$ is to cross-sectional unit $j$. This “similarity measure” $s_{ij}$ is large when the two units are similar, except that, for convenience but without loss of generality, we set $s_{ii} = 0$.

$s_i^+ \equiv \sum_j s_{ij}$ If $s_{ij}$ is zero or one for all $i$ and $j$, $s_i^+$ is known as the degree of $i$ and is interpreted as the number of $i$’s neighbors (or the number of edges connected to vertex $i$).

$\Sigma$ an unknown covariance matrix

$t$ a generic time period (usually a year)

$T$ total number of time periods (length of time series, when they all have the same length)

$T_i$ number of observations for cross section $i$ (if $T_i = T_j$, $\forall i, j = 1, \ldots, N$ then we set $T_i = T$)

$\theta$ drift parameter in the Lee-Carter model. We reuse this symbol for the smoothing parameter in our approach.

$U_{it}$ a missingness indicator equal to 0 if the dependent variable is missing in cross section $i$ at time $t$, and 1 if observed

$V$ generic orthogonal matrix

$V[\cdot]$ the variance

$W$ a symmetric matrix constructed from the similarity matrix $s$. See appendix B.2.6 (page 237).

$X_{it} \equiv U_{it} \sqrt{b_{it}} Z_{at}$ the explanatory variable vector ($X_{it}$) weighted by the exogenous weights $b_{it}$, when observed ($U_{it} = 1$) and 0 when missing

$\xi$ an error term

$x_{\circ}$ the projection of the vector $x$ on some subspace

$x_{\perp}$ the projection of the vector $x$ on the orthogonal complement of some subspace
APPENDIX A

\[ y_{it} \equiv U_{it} \sqrt{P_{it}} m_{it} \]

log-mortality rate \((m_{it})\) weighted by population \((p_{it})\), when observed \((U_{it} = 1)\) and 0 when missing

\(Z_{it}\)
a \(k_i\)-dimensional vector of covariates, for cross-sectional unit \(i\) at time \(t\). The vector of covariates usually includes the constant.

\(Z_i\)
the \(k_i \times T_j\) data matrix for cross section \(i\), whose rows are given by the vectors \(Z_{it}\)

\(Z\)
the set of integers