Appendixes

Appendix A

Notation

A.1 Principles

Variables and Parameters We use Greek symbols for *unknown quantities*, such as regression coefficients ($\boldsymbol{\beta}$), expected values (μ), disturbances (ϵ), and variances (σ^2), and Roman symbols for *observed quantities*, such as y and m for the dependent variable, while the symbols $\mathbf{X}$ and $\mathbf{Z}$ refer to covariates.

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

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 $\mathbf{Z}_{it}$ is a $k_i \times 1$ vector of covariates and $\boldsymbol{\beta}_i$ is a $k_i \times 1$ vector of coefficients. Every vector or matrix with one or more dimensions equal to k_i, such as $\mathbf{Z}_{it}$ or $\boldsymbol{\beta}_i$, 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_t 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 $\boldsymbol{\beta}$ the totality of vectors $\boldsymbol{\beta}_i$. Similarly, $\mathbf{Z}_i$ denotes the standard $T \times k_i$ data matrix for cross section i, with rows equal to the vector $\mathbf{Z}_{it}$.

If $\mathbf{X}$ is a vector, then diag[$\mathbf{X}$] is the diagonal matrix with $\mathbf{X}$ on its diagonal. If W is a matrix, then 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.

Norms For a matrix $\mathbf{x}$, we define the weighted Euclidean (or Mahalanobis) norm as $\|\mathbf{x}\|_\Phi^2 \equiv \mathbf{x}'\Phi\mathbf{x}$, 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 μ and standard deviation σ 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 $\mathcal{P}_{\mathbf{x}}(\mathbf{x})$ and $\mathcal{P}_{\mathbf{z}}(\mathbf{z})$ we simply write $\mathcal{P}(\mathbf{x})$ and $\mathcal{P}(\mathbf{z})$.

Sets Sets such as the real line $\mathbb{R}$ and its subsets ($\mathbb{S} \subset \mathbb{R}$) 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 $\mathfrak{N}$.

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
$\boldsymbol{\beta}_k^{\text{WLS}} \equiv (\mathbf{X}_k'\mathbf{X}_k)^{-1}\mathbf{X}_k' y_k$	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
δ_{ij}	Kronecker's delta function, equal to 1 if $i = j$ and 0 otherwise
$\text{E}[\cdot]$	the expected value operator
ϵ	an error term
$F(\mu)$	summary measures
η	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
λ	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.
$\bar{m}_a$	mean log-mortality age profile, averaging over time, $\sum_{t=1}^{T} m_{at}/T$

$\tilde{m}$	matrix of mean-centered logged mortality rates, with elements $\tilde{m}_{at} \equiv m_{at} - \frac{1}{T}\sum_t m_{at}$
μ_{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
$\mathfrak{n}$	generic order of the derivative of the smoothness functional
$\mathfrak{N}$	the null space of an operator or a functional
$\mathfrak{N}_\perp$	the orthogonal complement of the null space $\mathfrak{N}$
ν	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).
Σ	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 $T_i = T$)
θ	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
$\mathrm{V}[\cdot]$	the variance
W	a symmetric matrix constructed from the similarity matrix s. See appendix B.2.6 (page 237).
$\mathbf{X}_{it} \equiv U_{it}\sqrt{b_{it}}\mathbf{Z}_{it}$	the explanatory variable vector ($\mathbf{X}_{it}$) weighted by the exogenous weights b_{it} when observed ($U_{it} = 1$) and 0 when missing
ξ	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

$y_{it} \equiv U_{it}\sqrt{p_{it}}m_{it}$	log-mortality rate (m_{it}) weighted by population (p_{it}), when observed ($U_{it} = 1$) and 0 when missing
$\mathbf{Z}_{it}$	a k_i-dimensional vector of covariates, for cross-sectional unit i at time t. The vector of covariates usually includes the constant.
$\mathbf{Z}_i$	the $k_i \times T_i$ data matrix for cross section i, whose rows are given by the vectors $\mathbf{Z}_{it}$
$\mathbb{Z}$	the set of integers