Econometrics Journal (2002), volume **5**, pp. 76–90.

Notation in econometrics: a proposal for a standard

KARIM M. ABADIR[†‡] AND JAN R. MAGNUS[‡]

[†]*Departments of Mathematics and Economics, University of York, York, UK*
E-mail: kma4@york.ac.uk
[‡]*CentER, Tilburg University, Tilburg, The Netherlands*
E-mail: magnus@kub.nl

Received: January 2001

Summary This paper proposes a standard for notation in econometrics. It presents a fully integrated and internally consistent framework for notation and abbreviations, which is as close as possible to existing common practice. The symbols used are instantly recognizable and interpretable, thus minimizing ambiguity and enhancing reading efficiency. The standard is designed in a flexible manner, thus allowing for future extensions.

Keywords: *Notation, Symbols, Econometrics.*

1. INTRODUCTION

Few things are as boring as questions of notation. Serious researchers should do serious research and not waste their time thinking about notation. The mathematician Littlewood said about Jordan that if he (Jordan) had four things on the same footing (such as a, b, c, d) they would appear as

$$a, \quad M_3', \quad \varepsilon_2, \quad \Pi_{1,2}'';$$

see Bollobás (1986, p. 60).

On the other hand, notation—as we have discovered—raises passions, and many serious researchers *did* worry about notation. Jan Tinbergen proposed that 'when you have an index to a certain variable you should use the capital letter as its upper limit.' For example, $i = 1, \ldots, I$ and $j = 1, \ldots, J$, because this 'was just a little detail that could help you a lot to see through things' (Magnus and Morgan, 1987, p. 127).

In physics, engineering, and chemistry a serious attempt has been made to standardize symbols. The International Organization for Standardization (ISO) has published international regulations (ISO Standards Handbook, 1982) and the international union of pure and applied physics (IPU) has issued recommendations (CRC Handbook of Chemistry and Physics, 1988). These regulations are generally followed by the profession.

Notation matters. A good and consistent notation helps in the understanding, communication and development of our profession. If ill chosen, notation can be a complete barrier to comprehensibility. In the Renaissance, mathematics was written in a verbal style with p for plus, m for minus and R for square root. So, when Cardano (1501–1576) writes

5p : Rm : 15

5m : Rm : 15

25m : m : 15 qd est 40,

he means $(5+\sqrt{-15})(5-\sqrt{-15}) = 25-(-15) = 40$, see Kline (1972, p. 260). There is no doubt that the development of good notation has been of great importance in the history of mathematics.

In this paper we attempt to harmonize the various practices in econometrics notation. It proposes a fully integrated and internally consistent framework for notation and abbreviations, staying as close as possible to existing common practice. The symbols used are instantly recognizable and interpretable, thus minimizing ambiguity and enhancing reading efficiency. Using a common notation will save authors the effort to define their notation in every paper. Only special notation needs to be defined. We have tried to design our standard in a flexible manner, allowing for future extensions in specialized fields.

What we propose is a mix of standardization and reform. Our main goal is to reach agreement on basic notation, a certain minimum number of rules that econometricians can remember and follow painlessly.

There are many problems in designing a consistent notation. Our hope is to provide a useful benchmark and starting point for an evolving process. The notation is LaTeX oriented. Many new LaTeX commands are provided.[1] The complete list of LaTeX commands can be downloaded from

`http://cwis.kub.nl/~few5/center/staff/magnus/notation.htm`.

2. VECTORS AND MATRICES

Vectors are lowercase and matrices are uppercase symbols. Moreover, both vectors and matrices are written in bold-italic. The vectors $\boldsymbol{a}$, $\boldsymbol{b}, \ldots, \boldsymbol{z}$ are produced by `\va`, `\vb`, ..., `\vz`, and the matrices $\boldsymbol{A}$, $\boldsymbol{B}, \ldots, \boldsymbol{Z}$ by `\mA`, `\mB`, ..., `\mZ`.

Vectors can also be denoted by Greek lowercase letters: $\boldsymbol{\alpha}, \ldots, \boldsymbol{\omega}$ (`\valpha`, ..., `\vomega`), and matrices by Greek uppercase letters, such as $\boldsymbol{\Gamma}$ (`\mGamma`) or $\boldsymbol{\Theta}$ (`\mTheta`).

We write

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \qquad \boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}$$

for an $n \times 1$ vector $\boldsymbol{a}$ and an $m \times n$ matrix $\boldsymbol{A}$. If one has a choice, we recommend that the number of rows (m) exceeds or equals the number of columns (n).

We write $\boldsymbol{A} = (a_{ij})$ or $\boldsymbol{A} = (\boldsymbol{A})_{ij}$ to denote a typical element of the matrix $\boldsymbol{A}$. The n columns of $\boldsymbol{A}$ are denoted by $\boldsymbol{a}_{\bullet 1}, \boldsymbol{a}_{\bullet 2}, \ldots, \boldsymbol{a}_{\bullet n}$, and the m rows by $\boldsymbol{a}'_{1\bullet}, \boldsymbol{a}'_{2\bullet}, \ldots, \boldsymbol{a}'_{m\bullet}$, where transpose is denoted by a prime.[2] Hence,

$$\boldsymbol{A} = (\boldsymbol{a}_{\bullet 1}, \boldsymbol{a}_{\bullet 2}, \ldots, \boldsymbol{a}_{\bullet n}), \qquad \boldsymbol{A}' = (\boldsymbol{a}_{1\bullet}, \boldsymbol{a}_{2\bullet}, \ldots, \boldsymbol{a}_{m\bullet}).$$

However, we may write $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n)$, and occasionally $\boldsymbol{A}' = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m)$, when there is no possibility of confusion, but not both of these in one paper. A vector $\boldsymbol{a}$ denotes a column and $\boldsymbol{a}'$ denotes a row. Special vectors are:

[1] Only three LaTeX commands have been redefined: `\Re` which now reads Re, `\Im` which now reads Im, and `\vec` which now reads vec.

[2] The symbol $\cdot$ is produced by `\bcdot` since `\cdot` ($\cdot$) is too small and `\bullet` ($\bullet$) is too large.

$\mathbf{0}$, $\mathbf{0}_n$	null vector $(0, 0, \ldots, 0)'$ (\vzeros)
$\boldsymbol{\iota}$, $\boldsymbol{\iota}_n$	sum vector $(1, 1, \ldots, 1)'$ (\vones)
$\boldsymbol{e}_i$, $\boldsymbol{e}_{i,n}$	i-th column of $\boldsymbol{I}_n$ (\ve).

Special matrices are:

$\mathbf{O}$, $\mathbf{O}_{mn}$	null matrix of order $m \times n$ (\mZeros)
$\boldsymbol{I}$, $\boldsymbol{I}_n$	identity matrix of order $n \times n$ (\mI).

Note that the null vector $\mathbf{0}$ is smaller than the null matrix $\mathbf{O}$. We say that two or more matrices (vectors) are *conformable* if their sum or product is defined. For example, the equation $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ only makes sense if the order of $\boldsymbol{x}$ equals the number of columns of $\boldsymbol{A}$ and the order of $\boldsymbol{b}$ equals the number of its rows. If this is the case then $\boldsymbol{A}$, $\boldsymbol{x}$ and $\boldsymbol{b}$ are conformable.

3. OPERATIONS ON MATRIX A AND VECTOR a

Let $\boldsymbol{A}$ be a real $m \times n$ matrix of rank r. Then, $\boldsymbol{A}$ can be viewed as a collection of n columns in $\mathbb{R}^m$, but also as a collection of m rows in $\mathbb{R}^n$. Thus, associated with $\boldsymbol{A}$ are two vector spaces, the collection of columns and the collection of rows. In each of the two spaces there are two subspaces of special importance. The *column space* of $\boldsymbol{A}$, denoted col $\boldsymbol{A}$ or col($\boldsymbol{A}$) (\col), consists of all linear combinations of the columns,

$$\operatorname{col} \boldsymbol{A} = \{\boldsymbol{x} \in \mathbb{R}^m : \boldsymbol{x} = \boldsymbol{A}\boldsymbol{y} \text{ for some } \boldsymbol{y} \in \mathbb{R}^n\}.$$

The dimension (\dim) of col $\boldsymbol{A}$ is dim (col $\boldsymbol{A}$) $= r$. The *kernel* (or *null space*) of $\boldsymbol{A}$, denoted ker $\boldsymbol{A}$ or ker($\boldsymbol{A}$) (\ker), is the set ker $\boldsymbol{A} = \{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{y} = \mathbf{0}\}$ with dimension dim(ker $\boldsymbol{A}$) $= n-r$. The column space and kernel of $\boldsymbol{A}'$ are defined in the same way. The two kernels are more commonly known as *orthogonal complements* $\operatorname{col}^{\perp}(\boldsymbol{A}) = \ker \boldsymbol{A}'$ and $\operatorname{col}^{\perp}(\boldsymbol{A}') = \ker \boldsymbol{A}$.

Two vectors $\boldsymbol{x}$ and $\boldsymbol{a}$ for which $\boldsymbol{x}'\boldsymbol{a} = 0$ are *orthogonal*, and we write $\boldsymbol{x} \perp \boldsymbol{a}$ (\bot). If $\boldsymbol{x}$ is orthogonal to all columns of $\boldsymbol{A}$, we write $\boldsymbol{x} \perp \boldsymbol{A}$. Thus, $\operatorname{col}^{\perp}(\boldsymbol{A}) = \{\boldsymbol{x} : \boldsymbol{x} \perp \boldsymbol{A}\}$ with dimension dim($\operatorname{col}^{\perp} \boldsymbol{A}$) $= m - r$. A basis of $\operatorname{col}^{\perp}(\boldsymbol{A})$ is denoted $\boldsymbol{A}_{\perp}$. Hence, $\boldsymbol{A}_{\perp}$ denotes any $m \times (m-r)$ matrix with full column-rank, satisfying $\boldsymbol{A}'\boldsymbol{A}_{\perp} = \mathbf{O}$.

The following standard operations are proposed:

$\boldsymbol{A}'$	transpose
$\boldsymbol{A}^{-1}$	inverse
$\boldsymbol{A}^{+}$	Moore–Penrose inverse
$\boldsymbol{A}^{-}$	generalized inverse (satisfying only $\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{A} = \boldsymbol{A}$)
dg $\boldsymbol{A}$, dg($\boldsymbol{A}$)	diagonal matrix containing the diagonal elements of $\boldsymbol{A}$ (\dg)
diag($a_1, \ldots, a_n$)	diagonal matrix containing $a_1, \ldots, a_n$ on the diagonal (\diag)
diag($\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$)	block-diagonal matrix with $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$ on the diagonal
$\boldsymbol{A}^2$	$\boldsymbol{A}\boldsymbol{A}$
$\boldsymbol{A}^{1/2}$	(unique) square root of positive semidefinite matrix
$\boldsymbol{A}^{p}$	pth power
$\boldsymbol{A}^{\#}$	adjoint (matrix)
$\boldsymbol{A}^{*}$	conjugate transpose, $\boldsymbol{A}^{*} = \boldsymbol{U}' - \mathrm{i}\boldsymbol{V}'$ when $\boldsymbol{A} = \boldsymbol{U} + \mathrm{i}\boldsymbol{V}$, $\boldsymbol{U}$ and $\boldsymbol{V}$ real, and $\mathrm{i} = \sqrt{-1}$

$\boldsymbol{A}_k$	principal submatrix of order $k \times k$
$(\boldsymbol{A}, \boldsymbol{B})$, $(\boldsymbol{A} : \boldsymbol{B})$	partitioned matrix
vec $\boldsymbol{A}$, vec$(\boldsymbol{A})$	vec operator (\vec)
vech $\boldsymbol{A}$, vech$(\boldsymbol{A})$	vector containing a_{ij} $(i \geq j)$ (\vech)
rk$(\boldsymbol{A})$	rank (\rk)
λ_i, $\lambda_i(\boldsymbol{A})$	ith eigenvalue (of $\boldsymbol{A}$)
tr $\boldsymbol{A}$, tr$(\boldsymbol{A})$	trace (\tr)
etr $\boldsymbol{A}$, etr$(\boldsymbol{A})$	exp(tr $\boldsymbol{A}$) (\etr)
$\lvert\boldsymbol{A}\rvert$, det $\boldsymbol{A}$, det$(\boldsymbol{A})$	determinant (\det)
$\lVert\boldsymbol{A}\rVert$	norm of matrix $(\sqrt{(\text{tr}\,\boldsymbol{A}^*\boldsymbol{A})})$ (\\|)
$\lVert\boldsymbol{a}\rVert$	norm of vector $(\sqrt{(\boldsymbol{a}^*\boldsymbol{a})})$
$\boldsymbol{A} \geq \boldsymbol{B}$, $\boldsymbol{B} \leq \boldsymbol{A}$	$\boldsymbol{A} - \boldsymbol{B}$ positive semidefinite (\geq\ge, \leq\le)
$\boldsymbol{A} > \boldsymbol{B}$, $\boldsymbol{B} < \boldsymbol{A}$	$\boldsymbol{A} - \boldsymbol{B}$ positive definite ($>$, $<$)
$\boldsymbol{A} \otimes \boldsymbol{B}$	Kronecker product (\otimes)
$\boldsymbol{A} \odot \boldsymbol{B}$	Hadamard product (\odot).

A few words of explanation on some of the symbols is required. First, the square root of a positive semidefinite matrix $\boldsymbol{A} = \boldsymbol{S}\boldsymbol{\Lambda}\boldsymbol{S}'$ ($\boldsymbol{S}$ orthogonal, $\boldsymbol{\Lambda}$ diagonal) is defined here as the unique matrix $\boldsymbol{A}^{1/2} = \boldsymbol{S}\boldsymbol{\Lambda}^{1/2}\boldsymbol{S}'$. Next, ambiguity can arise between the symbol $|\cdot|$ for determinant and the same symbol for absolute value, for example in the multivariate transformation theorem. This ambiguity can be avoided by writing $|\det A|$ for the absolute value of a determinant. Finally, in mathematical economics we encounter positive (non-negative) matrices, that is, matrices with the property that all elements are positive (non-negative). Positivity can be denoted by $\boldsymbol{A} \succ \boldsymbol{O}$ (\succ) and non-negativity by $\boldsymbol{A} \succeq \boldsymbol{O}$ (\succeq).

In addition, we propose standards for the following matrices:

$\boldsymbol{K}_{mn}$	commutation matrix
$\boldsymbol{K}_n$	$\boldsymbol{K}_{nn}$
$\boldsymbol{N}_n$	$\frac{1}{2}(\boldsymbol{I}_{n^2} + \boldsymbol{K}_n)$
$\boldsymbol{D}_n$	duplication matrix
$\boldsymbol{J}_k(\lambda)$	Jordan block of order $k \times k$.

If an $n \times n$ matrix $\boldsymbol{A}$ is symmetric, then its eigenvalues are real and can be ordered. We recommend the ordering

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n,$$

since there are many cases (e.g. principal-components analysis) where the largest eigenvalues are the ones of interest.

4. GREEK SYMBOLS

Some Greek lowercase letters have variant forms and these can be used to mean different things than the usual letter. We have:

ϵ	\epsilon, \eps	ε	\varepsilon, \epsi
θ	\theta	ϑ	\vartheta
π	\pi	ϖ	\varpi
ρ	\rho	ϱ	\varrho

σ	\sigma	ς	\varsigma
ϕ	\phi	φ	\varphi.

We shall use ε (\epsi for a scalar, \vepsi for a vector) for a disturbance term and ϵ (\eps) for an arbitrarily small positive number. Also, we use θ (\theta) to denote a variable and ϑ (\vartheta) for a function.

5. MATHEMATICAL SYMBOLS, FUNCTIONS AND OPERATORS

Definitions, implications, convergence, and transformations are denoted by:

$\equiv$	identity, equivalence (\equiv)
$a := b$	defines a in terms of b
$\Longrightarrow$	implies (\implies)
$\Longleftrightarrow$	if and only if (\iff)
$\to, \longrightarrow$	converges to (\to, \longto)
$x \mapsto y$	transformation from x to y (\mapsto).

We write $f(x) \approx g(x)$ (\approx) if the two functions are approximately equal in some sense depending on the context. If $f(x)$ is proportional to $g(x)$ we write $f(x) \propto g(x)$ (\propto). We say that '$f(x)$ is at most of order $g(x)$' and write $f(x) = O(g(x))$, if $|f(x)/g(x)|$ is bounded from above in some neighborhood of $x = c$ (possibly $c = \pm\infty$), and we say that '$f(x)$ is of order less than $g(x)$' and write $f(x) = o(g(x))$, if $f(x)/g(x) \to 0$ when $x \to c$. Finally, we write $f(x) \sim g(x)$ (\sim) if $f(x)/g(x) \to 1$ when $x \to c$. The two functions are then said to be 'asymptotically equal'.[3] Notice that when $f(x)$ and $g(x)$ are asymptotically equal, then $f(x) \approx g(x)$ and also $f(x) = O(g(x))$, but not vice versa.

For example, when ϕ and Φ denote the standard-normal density and distribution functions, respectively, we write the leading term (first term) of the asymptotic expansion

$$\frac{\Phi(x)}{\phi(x)} \sim \frac{1}{|x|} \qquad \text{as } x \to -\infty.$$

However, there are many good local approximations of this ratio which are not necessarily asymptotically equal to it.

The usual sets are denoted as follows:

$\mathbb{N}$	natural numbers $1, 2, \ldots$	\SN
$\mathbb{Z}$	integers $\ldots, -2, -1, 0, 1, 2, \ldots$	\SZ
$\mathbb{Q}$	rational numbers	\SQ
$\mathbb{R}$	real numbers	\SR
$\mathbb{C}$	complex numbers	\SC.

Superscripts denote the dimension and subscripts the relevant subset. For example, $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ denotes the real plane, $\mathbb{R}^n$ the set of real $n \times 1$ vectors, and $\mathbb{R}^{m \times n}$ the set of real $m \times n$ matrices. The set $\mathbb{R}^n_+$ denotes the positive orthant of $\mathbb{R}^n$, while $\mathbb{Z}_+$ denotes the set of positive integers

[3] The ISO prescribes the symbol $\simeq$ (\simeq) for asymptotic equality, but $\sim$ is common practice in econometrics and statistics, even though the same symbol is also used for 'is distributed as'.

(hence, $\mathbb{Z}_+ = \mathbb{N}$) and $\mathbb{Z}_{0,+}$ denotes the non-negative integers. Finally, $\mathbb{C}^{n\times n}$ denotes the set of complex $n \times n$ matrices.

Set differences are denoted by a backslash (`\backslash`). For example, $\mathbb{N} = \mathbb{Z}_{0,+}\backslash\{0\}$. Real-line intervals defined for x by $a \leq x < b$ are denoted by $[a, b)$. Occasionally it might be unclear whether (a, b) indicates a real-line interval or a point in $\mathbb{R}^2$. In that case the interval $a < x < b$ can alternatively be written as $]a, b[$.

Sequences are special ordered sets. They are delimited, as usual for sets, by braces (curly brackets). It is often convenient to write $\{\boldsymbol{Z}_j\}_m^n$ (or simply $\{\boldsymbol{Z}_j\}$) for the sequence of matrices $\boldsymbol{Z}_m, \boldsymbol{Z}_{m+1}, \ldots, \boldsymbol{Z}_n$.

Other symbols used are:

$\in$	belongs to (`\in`)
$\notin$	does not belong to (`\notin`)
$\{x : x \in S, x \text{ satisfies } P\}$	set of all elements of S with property P
$\subseteq$	is a subset of (`\subseteq`)
$\subset$	is a proper subset of (`\subset`)
$\cup$	union (`\cup`)
$\cap$	intersection (`\cap`)
$\emptyset$	empty set (`\emptyset`)
A^c	complement of A
$B\backslash A$	$B \cap A^c$.

We denote functions by:

$f : S \to T$	function defined on S with values in T
$f, g, \varphi, \psi, \vartheta$	scalar-valued function
$\boldsymbol{f}, \boldsymbol{g}$	vector-valued function
$\boldsymbol{F}, \boldsymbol{G}$	matrix-valued function
$\boldsymbol{g} \circ \boldsymbol{f}, \boldsymbol{G} \circ \boldsymbol{F}$	composite function (`\circ`)
$g * f$	convolution $(g * f)(x) = \int_{-\infty}^{\infty} g(y) f(x - y)\,\mathrm{d}y$.

The next issue is controversial. While the ISO regulations are generally followed by the profession, there is one major exception: the treatment of lowercase single-letter constants. Examples are the base of natural logarithms e and the imaginary unit i. These are often written as e and i, contrary to ISO regulations. The same applies to operators (such as the derivative operator d—often written as d).[4] It appears that the profession finds that single-letter lowercase roman mathematical symbols look odd. There are examples of this phenomenon in econometrics too: one often sees $\det(A)$ for determinant, $\mathrm{E}(x)$ for expectation, but $r(A)$ for rank. We follow the ISO instructions and recommend i, e, and d. We believe there are good reasons for doing this. The imaginary unit i should not be confused with the index i. Also, since log is in roman type it would be odd to write e instead of e. Finally, since essentially all operators are in roman type, why not write d instead of d for the differential, thereby also avoiding possible confusion with many other uses of d such as distance.

Thus, for differentials, derivatives and differences, we write:

d	differential (`\rd`)
d^n	nth order differential

[4] See Beccari (1997) for further discussion and some LaTeX tricks for physicists and engineers.

$\mathrm{D}_j\varphi(\boldsymbol{x})$	partial derivative (\rD), $\partial\varphi(\boldsymbol{x})/\partial x_j$
$\mathrm{D}_j f_i(\boldsymbol{x})$	partial derivative, $\partial f_i(\boldsymbol{x})/\partial x_j$
$\mathrm{D}^2_{kj}\varphi(\boldsymbol{x})$	second-order partial derivative, $\partial\mathrm{D}_j\varphi(\boldsymbol{x})/\partial x_k$
$\mathrm{D}^2_{kj} f_i(\boldsymbol{x})$	second-order partial derivative, $\partial\mathrm{D}_j f_i(\boldsymbol{x})/\partial x_k$
$\varphi^{(n)}(x)$	nth order derivative of $\varphi(x)$
$\mathrm{D}\varphi(\boldsymbol{x})$, $\partial\varphi(\boldsymbol{x})/\partial\boldsymbol{x}'$	derivative of $\varphi(\boldsymbol{x})$
$\mathrm{D}\boldsymbol{f}(\boldsymbol{x})$, $\partial\boldsymbol{f}(\boldsymbol{x})/\partial\boldsymbol{x}'$	derivative (Jacobian matrix) of $\boldsymbol{f}(\boldsymbol{x})$
$\mathrm{D}\boldsymbol{F}(\boldsymbol{X})$	derivative (Jacobian matrix) of $\boldsymbol{F}(\boldsymbol{X})$
$\partial\,\mathrm{vec}\,\boldsymbol{F}(\boldsymbol{X})/\partial(\mathrm{vec}\,\boldsymbol{X})'$	derivative of $\boldsymbol{F}(\boldsymbol{X})$, alternative notation
$\nabla\varphi$, $\nabla\boldsymbol{f}$, $\nabla\boldsymbol{F}$	gradient (transpose of derivative) (\nabla)
$\mathrm{H}\varphi(\boldsymbol{x})$, $\partial^2\varphi(\boldsymbol{x})/\partial\boldsymbol{x}\partial\boldsymbol{x}'$	second derivative (Hessian matrix) of $\varphi(\boldsymbol{x})$ (\rH)
L, B	backward shift operator: $\mathrm{L}x_t = x_{t-1}$ (\rL,\rB)
Δ, Δ_b	(backward) difference operator: $\Delta x_t = x_t - x_{t-1}$ (\diff, \bdiff)
Δ_f	forward difference operator: $\Delta_\mathrm{f}x_t = x_{t+1} - x_t$ (\fdiff)
Δ^s	$(1-\mathrm{L})^s$
Δ_s	$1-\mathrm{L}^s$
$[f(x)]_a^b$, $f(x)\|_a^b$	$f(b) - f(a)$.

Instead of $\varphi^{(1)}(x)$ and $\varphi^{(2)}(x)$, one can write the more common $\varphi'(x)$ and $\varphi''(x)$, but otherwise we prefer to reserve the prime for matrix transposes rather than derivatives. To emphasize the difference between transpose and derivative, we write $\boldsymbol{f}'(\boldsymbol{x})$ for the derivative of $\boldsymbol{f}$ and $\boldsymbol{f}(\boldsymbol{x})'$ for the transpose. The distinction between backward and forward differencing is actually important. In the appendix we illustrate how bad notation can lead to incorrect results in this area.

We use L (or B) for the lag operator rather than $\mathcal{L}$ (Laplace transform) or L (likelihood), in order to avoid confusion. Some useful transforms are:

$\mathcal{F}\{\cdot\}$	Fourier transform (\calF)
$\mathcal{F}^{-1}\{\cdot\}$	inverse Fourier transform
$\mathcal{L}\{\cdot\}$	Laplace transform (\calL)
$\mathcal{L}^{-1}\{\cdot\}$	inverse Laplace transform.

Finally, various other symbols in common use are:

i	imaginary unit (\iu)
e, exp	exponential (\eu, \exp)
log	natural logarithm (\log)
$\log_a$	logarithm to the base a
!	factorial
δ_{ij}	Kronecker delta
$\mathrm{sgn}(x)$	sign of x (\sgn)
$\lfloor x \rfloor$, $\mathrm{int}(x)$	integer part of x, that is, largest integer $\leq x$ (\lfloor, \rfloor, \ip)
$\|x\|$	absolute value (modulus) of scalar $x \in \mathbb{C}$
x^*	complex conjugate of scalar $x \in \mathbb{C}$
$\mathrm{Re}(x)$	real part of x (\Re)
$\mathrm{Im}(x)$	imaginary part of x (\Im)

$\Gamma(x)$	gamma (generalized factorial) function, satisfying $\Gamma(x+1) = x\Gamma(x)$
$B(x, y)$	beta function, $\Gamma(x)\Gamma(y)/\Gamma(x+y)$
$1_{\mathcal{K}}$	indicator function (use 1, not I): equals 1 if condition $\mathcal{K}$ is satisfied, 0 otherwise
$B(\boldsymbol{c})$, $B(\boldsymbol{c}; r)$	neighborhood (ball) with center $\boldsymbol{c}$ and radius r
$\mathcal{V}^{n\times k}$	Stiefel manifold: set of real $n \times k$ matrices $\boldsymbol{X}$ such that $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{I}_k$ $(k \leq n)$ (\calV)
$\mathcal{O}^n$	$\mathcal{V}^{n\times n}$, orthogonal group of dimension n (\calO)
$\mathcal{O}^n_+$	proper orthogonal group of dimension n (orthogonal $n \times n$ matrices with determinant $+1$)
$\mathcal{S}^n$	$\mathcal{V}^{n\times 1}$, unit sphere in $\mathbb{R}^n$ (\calS).

The Stiefel manifold $\mathcal{V}^{n\times k}$ is also denoted as $\mathcal{V}^{k\times n}$ in the literature. We recommend the former notation which is in line with $\mathbb{R}^{n\times k}$.

6. STATISTICAL SYMBOLS, FUNCTIONS AND OPERATORS

The final major problem of notation is the statistical convention to write capital letters for random variables and lowercase letters for their realizations, for example $\Pr(X = x)$. In the multivariate case, we cannot use uppercase for random variables (e.g. $\boldsymbol{X}$) and lowercase for their realizations (e.g. $\boldsymbol{x}$), because this leads to an inconsistency in notation: vectors should be lowercase and matrices uppercase. The same problem would arise in the scalar case, since one would run into an inconsistency in choosing a label for the random vector (X, Y, Z) and its realization (x, y, z). This is a well-known problem and all authors of econometrics textbooks have to find some solution to it. There is no solution which is completely satisfactory. Perhaps the old Dutch custom to underline random variables and write for example $\Pr(\underline{x} = x)$ and $\mathrm{E}(\underline{x}|\underline{y} = y)$ is still the best. But we have no hope that this convention will be followed.

Thus we propose the following. If there are up to three random variables, we write (x, y, z) for the random variables and (u, v, w) for their realizations. For example, $f_{x,y}(u, v)$ for the joint density of x and y, and $\mathrm{E}(x|y = v)$ for the conditional expectation of x given $y = v$. If we have a vector $\boldsymbol{x}$ of random variables, we write $f_{\boldsymbol{x}}(\boldsymbol{w})$. (We don't use $\boldsymbol{u}$ and $\boldsymbol{v}$ for realizations because they are already in use for error terms in multivariate models.) If we need up to three vectors, we can write these as $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ with realizations $\boldsymbol{w}_{\boldsymbol{x}}$, $\boldsymbol{w}_{\boldsymbol{y}}$ and $\boldsymbol{w}_{\boldsymbol{z}}$. For more than three, we use $\boldsymbol{x}_j$ $(j = 1, \ldots, m)$ with realizations $\boldsymbol{w}_j$, and marginal densities $f_j(\boldsymbol{w}_j)$. Similarly, in the matrix case, $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are matrices of random variables and their realizations are $\boldsymbol{W}_{\boldsymbol{X}}$, $\boldsymbol{W}_{\boldsymbol{Y}}$ and $\boldsymbol{W}_{\boldsymbol{Z}}$; and so on for the case of $m > 3$. The subscripts are only used when there is a possibility of confusion. This convention appears to work well in practice and is consistent with econometric notation.

We follow the standard convention to denote by F the cumulative distribution function (c.d.f.), and by f the probability density function (p.d.f.). These will in general depend on a vector of m parameters $\boldsymbol{\theta}$. An estimator of $\boldsymbol{\theta}$ is denoted $\hat{\boldsymbol{\theta}}$ and, if there is a second estimator, by $\tilde{\boldsymbol{\theta}}$. The realization of an estimator is an estimate. Predictors are like estimators, except that they say something about a random variable. They are also denoted by 'hats' $(\hat{\boldsymbol{y}}, \hat{\boldsymbol{\varepsilon}})$ or tildes $(\tilde{\boldsymbol{y}}, \tilde{\boldsymbol{\varepsilon}})$. The realization of a predictor is called a prediction.

We denote a null hypothesis as H_0 (\rH) and alternatives as $\mathrm{H}_1, \mathrm{H}_2, \ldots$. The value of $\boldsymbol{\theta}$ under hypothesis H_j is denoted by $\boldsymbol{\theta}^{(j)}$. If the hypothesis concerns only a subset $\boldsymbol{\theta}_i$, we denote the value under hypothesis H_j by $\boldsymbol{\theta}_i^{(j)}$.

If there exist r restrictions on $\boldsymbol{\theta}$ we write these as $\boldsymbol{h}(\boldsymbol{\theta}) = \mathbf{0}$ or, in the case of one restriction, as $h(\boldsymbol{\theta}) = 0$. The derivative of $\boldsymbol{h}$ is an $r \times m$ matrix $\boldsymbol{R}(\boldsymbol{\theta})' := \partial \boldsymbol{h}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'$, specializing to a $1 \times m$ row-vector $\boldsymbol{r}'$ in the case $r = 1$. The notation is chosen so that it corresponds to the linear case where we write the constraint as $\boldsymbol{R}'\boldsymbol{\theta} = \boldsymbol{c}$.

The following symbols are commonly used:

$\sim$	is distributed as (\distr)
$\overset{a}{\sim}$	is asymptotically distributed as (\adistr)
$\Pr(\cdot)$	probability (\Pr)
$\mathrm{E}(\cdot)$	expectation (\E)
$\mathrm{E}(\cdot\|\cdot)$, $\mathrm{E}_{x\|y}(\cdot)$	conditional expectation
$\mathrm{var}(\cdot)$	variance (matrix) (\var)
$\mathrm{cov}(\cdot,\cdot)$	covariance (matrix) (\cov)
$\mathrm{corr}(\cdot,\cdot)$	correlation (matrix) (\corr)
$L(\boldsymbol{\theta})$	likelihood function
$\ell(\boldsymbol{\theta})$	log-likelihood function (\ell)
$\boldsymbol{q}(\boldsymbol{\theta})$	score vector
$\boldsymbol{\mathcal{H}}(\boldsymbol{\theta})$	Hessian matrix (\Hesmat)
$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})$	(Fisher) information matrix (\Infmat)
$\mathcal{F}_t$	filtration at time t (\calF)
$\to$, $\longrightarrow$	converges a.s. (\to, \longto)
$\xrightarrow{p}$	converges in probability (\pto)
$\xrightarrow{d}$	converges in distribution (\dto)
$\xrightarrow{w}$	converges weakly (\wto)
plim	probability limit (\plim)
$O_p(g(x))$	at most of probabilistic order $g(x)$
$o_p(g(x))$	of probabilistic order less than $g(x)$.

Notice that the symbol $\to$ ($\longrightarrow$) indicates both convergence and a.s. convergence. The symbol $\xrightarrow{w}$ for weak convergence is preferred to $\Longrightarrow$, which denotes logical implication. When $\hat{\boldsymbol{\theta}}$ denotes the ML estimator, the matrix $-\boldsymbol{\mathcal{H}}(\hat{\boldsymbol{\theta}})$ is also called the observed information matrix. The matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) := -\mathrm{E}\left(\boldsymbol{\mathcal{H}}(\boldsymbol{\theta})\right)$ is the expected information matrix. When there is only one parameter ($m = 1$), we write $q(\theta)$ for the score, $\mathcal{H}(\theta)$ for the Hessian (\calH), and $\mathcal{I}(\theta)$ for the Fisher information (\calI).

The most common continuous distributions in econometrics are denoted as follows:

$\phi(\cdot)$	standard-normal p.d.f. (\phi)
$\Phi(\cdot)$	standard-normal c.d.f. (\Phi)
$\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Omega})$, $\mathrm{N}_m(\boldsymbol{\mu}, \boldsymbol{\Omega})$	m-dimensional normal distribution (\rN)
$\mathrm{IN}(\boldsymbol{\mu}_i, \boldsymbol{\Omega}_i)$, $\mathrm{IN}_m(\boldsymbol{\mu}_i, \boldsymbol{\Omega}_i)$	sequence $i = 1, 2, \ldots$ of independent m-dimensional normal distributions (\IN)
$\mathrm{LN}(\mu, \sigma^2)$	log-normal distribution (\LN)
$\chi^2(n, \delta)$	chi-squared distribution with n d.f. and non-centrality parameter δ

$\chi^2(n)$	central chi-squared ($\delta = 0$)
$\mathrm{t}(n, \delta)$	Student distribution with n d.f. and non-centrality δ (\rt)
$\mathrm{t}(n)$	central t ($\delta = 0$)
$\mathrm{F}(m, n, \delta)$	Fisher distribution with m (numerator) and n (denominator) d.f. and non-centrality δ (\rF)
$\mathrm{F}(m, n)$	central F ($\delta = 0$)
$\mathrm{Beta}_{(\alpha,\beta)}(p, q)$	beta distribution defined over the interval (α, β), with density function $\frac{(x-\alpha)^{p-1}(\beta-x)^{q-1}}{B(p,q)(\beta-\alpha)^{p+q-1}}$, $p, q > 0$ (\rBeta)
$\mathrm{Beta}(p, q)$	standard beta distribution, $\mathrm{Beta}_{(0,1)}(p, q)$
$\mathrm{U}_{(\alpha,\beta)}$, U	Uniform on (α, β), standard uniform on (0,1) (\rU); $\mathrm{U}_{(\alpha,\beta)} \equiv \mathrm{Beta}_{(\alpha,\beta)}(1, 1)$
$\mathrm{Gam}(\nu, \lambda)$	gamma distribution, with density function $\frac{\lambda^\nu x^{\nu-1}}{\Gamma(\nu)} \exp(-\lambda x)$, $\nu, \lambda > 0$ (\rGam).

In addition, we denote by $W(s)$ or $B(s)$ the standard Wiener process (Brownian motion) on $s \in [0, 1]$.

Quantiles are denoted as follows. If x follows some distribution, say $\mathrm{D}(\boldsymbol{\theta})$, then the αth quantile is $\mathrm{D}_\alpha(\boldsymbol{\theta})$. For example, $\mathrm{t}_{0.05}(n, \delta)$ denotes the 5% quantile of the non-central t-distribution. See also Section 8 for the notation $\mathrm{IID}_m(\boldsymbol{\theta})$ where m is a natural number, unlike α.

We use the word 'expectation' to denote mathematical expectation of a random vector $\boldsymbol{x}$, written $\mathrm{E}(\boldsymbol{x})$. The word 'average' refers to taking the average of some numbers: $\bar{x} = (1/n)\sum_{i=1}^n x_i$. The word 'mean', which can indicate either, is therefore ambiguous. Like 'expectation', the words 'variance' (var), 'covariance' (cov), and 'correlation' (corr) indicate population parameters. The corresponding sample parameters are called 'sample variance' ($\widehat{\mathrm{var}}$), 'sample covariance' ($\widehat{\mathrm{cov}}$), and 'sample correlation' ($\widehat{\mathrm{corr}}$).

Let $\hat{\theta}$ be an estimator of a parameter of interest θ. Suppose that $\sigma^2 = \mathrm{var}(\hat{\theta})$. Then, σ (the positive square root of the variance) is called the 'standard deviation' of $\hat{\theta}$. In general, σ is not known or depends on unknown parameters. Let $\hat{\sigma}$ be an estimator of σ. Its realization (the estimate) is called the 'standard error' of $\hat{\theta}$. The t-statistic (for significance from 0) is defined as $\hat{\theta}/\hat{\sigma}$, and is a random variable, not necessarily Student distributed. Its realization is the t-value.

7. THE LINEAR REGRESSION MODEL

We write the linear regression model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ as

$$\boldsymbol{y} = \sum_{h=1}^{k} \beta_h \boldsymbol{x}_{\bullet h} + \boldsymbol{\varepsilon}$$

or as

$$y_i = \boldsymbol{x}'_{i\bullet}\boldsymbol{\beta} + \varepsilon_i \quad (i = 1, 2, \ldots, n)$$

or as

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \quad (i = 1, 2, \ldots, n).$$

If there is a constant term this specializes to

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \quad (i = 1, 2, \ldots, n).$$

In the two-variable case one can write

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \qquad \text{or} \qquad y_i = \alpha + \beta x_i + \varepsilon_i,$$

but *not* $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, since β_0 is often used for other purposes, in particular as the value of the parameter β under the null hypothesis.

The observations are typically indexed $i = 1, \ldots, n$ (in cross sections) or $t = 1, \ldots, T$ (in time series). If there are two cross sections one can use i and j; if there are two time series one uses t and s. There are k regressors (not K) indexed by $h = 1, \ldots, k$. Acronyms and special symbols take precedence over index labels. For example, in defining the t-statistic one should not use t as a summation index.

This formulation is not without controversy. Some authors write x_{hi} instead of x_{ih}, which is unsatisfactory, since $\boldsymbol{X}$ is an $n \times k$ matrix and hence in their formulation x_{hi} is the ihth element of $\boldsymbol{X}$. Some write β_0 for the first element of $\boldsymbol{\beta}$, if the regression contains a constant term, and then let k denote the number of 'real' regressors (so that $\boldsymbol{X}$ has $k+1$ columns). We prefer to avoid this formulation for many reasons. It is convenient to always have k regressors independent of whether there is a constant term or not. Also, the inclusion of a constant does make an important difference, for example in potentially non-stationary time series, and it can translate into a 'real' variable such as a drift, which alters distributions and time paths.

Another issue is the disturbance term. We denote this by $\boldsymbol{\varepsilon}$ (`\epsi` for a scalar, `\vepsi` for a vector) if the disturbances (or errors) are spherically distributed.[5] If the errors are not spherical, we denote them by $\boldsymbol{u}$.

The symbols R^2 and $\bar{R}^2$ denote the coefficient of determination and the adjusted coefficient of determination, respectively.

In the case of OLS (ordinary least squares), it is tradition to write $\boldsymbol{b}$ instead of $\hat{\boldsymbol{\beta}}$ for the OLS estimator $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, $\boldsymbol{e}$ instead of $\hat{\boldsymbol{\varepsilon}}$ for the residuals, and s^2 instead of $\hat{\sigma}^2$ for the OLS estimator of σ^2.[6] We prefer not to do so, in order to stress the randomness of the estimators (one often thinks of $\boldsymbol{b}$ as a vector of constants and of $\boldsymbol{e}$ as a unit vector).

It is customary to write

$$\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{+}\boldsymbol{X}', \qquad \boldsymbol{M}_{\boldsymbol{X}} = \boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}},$$

when $\boldsymbol{X}$ has n rows. If there is no possibility of confusion, we can write $\boldsymbol{M}$ and $\boldsymbol{P}$ instead of $\boldsymbol{M}_{\boldsymbol{X}}$ and $\boldsymbol{P}_{\boldsymbol{X}}$. The matrix which puts a vector in deviation form is thus

$$\boldsymbol{M}_{\boldsymbol{\iota}} = \boldsymbol{I}_n - (1/n)\boldsymbol{\iota}\boldsymbol{\iota}',$$

and the vector $\boldsymbol{M}_{\boldsymbol{\iota}}\boldsymbol{a}$ denotes the vector $\boldsymbol{a}$ in deviation from its mean.

The statement of H_0: $\boldsymbol{R}'\boldsymbol{\beta} = \boldsymbol{c}$ is preferred over $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$.[7] In the latter formulation, the single-hypothesis case is usually written as $\boldsymbol{w}'\boldsymbol{\beta} = r$ or $\boldsymbol{r}'\boldsymbol{\beta} = r$, neither of which is ideal. However, if one writes $\boldsymbol{R}'\boldsymbol{\beta} = \boldsymbol{c}$, this specializes to $\boldsymbol{r}'\boldsymbol{\beta} = c$ in the one-dimensional case. This has the additional advantage that we can use r to denote the number of restrictions (order of $\boldsymbol{c}$).

[5] The vector $\boldsymbol{\varepsilon}$ is spherically distributed if $\boldsymbol{\varepsilon}$ and $\boldsymbol{S}\boldsymbol{\varepsilon}$ have identical distributions for every orthogonal matrix $\boldsymbol{S}$. The definition implies that if $\boldsymbol{\varepsilon}$ is spherical, then $c\boldsymbol{\varepsilon}$ is also spherical ($c \neq 0$).

[6] In line with current practice, we write the estimator for σ^2 as $\hat{\sigma}^2$ (`\hat{\sigma}^2`) and not as $\widehat{\sigma^2}$, although strictly speaking the latter is the correct notation.

[7] Our preferred notation implies that the null hypothesis $\boldsymbol{R}'\boldsymbol{\beta} = \boldsymbol{0}$ can also be written as $\boldsymbol{\beta} \perp \boldsymbol{R}$, or as $\boldsymbol{\beta} \in \text{col}^{\perp}(\boldsymbol{R})$, or as $\boldsymbol{\beta} = \boldsymbol{R}_{\perp}\boldsymbol{\alpha}$ for some $\boldsymbol{\alpha}$.

In the special case where $\boldsymbol{R} = \boldsymbol{I}_r$ (or where $\boldsymbol{R}$ is square and invertible), we usually write $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ rather than $\boldsymbol{\beta} = \boldsymbol{c}$.

The GLS model is written

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \qquad \boldsymbol{u} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Omega}).$$

If the variance matrix $\boldsymbol{\Omega}$ contains parameters, we write $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\alpha})$. The order of $\boldsymbol{\alpha}$ is l. Thus, the whole model contains $m = k + l$ unknown parameters. The complete parameter vector is denoted $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ and is of order m.

For the *simultaneous equations model* our starting point is the (univariate) linear regression model

$$y_i = \boldsymbol{x}_{i\bullet}'\boldsymbol{\beta} + u_i \quad (i = 1, 2, \ldots, n).$$

This can be generalized to the *multivariate* linear regression model:

$$\boldsymbol{y}_{i\bullet}' = \boldsymbol{x}_{i\bullet}'\boldsymbol{B} + \boldsymbol{u}_{i\bullet}' \quad (i = 1, 2, \ldots, n),$$

where $\boldsymbol{y}_{i\bullet}$ and $\boldsymbol{u}_{i\bullet}$ are random $m \times 1$ vectors and $\boldsymbol{B}$ is a $k \times m$ matrix. The univariate case is obtained as a special case when $m = 1$. The simultaneous equations model provides a further generalization:

$$\boldsymbol{y}_{i\bullet}'\boldsymbol{\Gamma} = \boldsymbol{x}_{i\bullet}'\boldsymbol{B} + \boldsymbol{u}_{i\bullet}' \quad (i = 1, 2, \ldots, n),$$

where $\boldsymbol{\Gamma}$ is an $m \times m$ matrix. This is the *structural form* of the simultaneous equations model. In matrix notation this becomes $\boldsymbol{Y}\boldsymbol{\Gamma} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{U}$. If $\boldsymbol{\Gamma}$ is invertible, we obtain the *reduced form* $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\Pi} + \boldsymbol{V}$, where $\boldsymbol{\Pi} = \boldsymbol{B}\boldsymbol{\Gamma}^{-1}$ and $\boldsymbol{V} = \boldsymbol{U}\boldsymbol{\Gamma}^{-1}$.

8. ABBREVIATIONS AND ACRONYMS

2SLS	two-stage least squares
3SLS	three-stage least squares
AIC	Akaike information criterion
AR(p)	autoregressive process of order p
ARCH	autoregressive conditional heteroskedasticity
ARIMA(p, d, q)	autoregressive integrated moving-average process
ARMA(p, q)	autoregressive moving-average process
a.s.	almost surely
BAN	best asymptotically normal
BIC	(Schwarz) Bayesian information criterion
B[L]U[E]	best [linear] unbiased [estimator]
c.d.f.	cumulative distribution function
c.f.	characteristic function
c.g.f.	cumulant-generating function
CLS	constrained least squares (‘restricted’ LS)
CLT	central limit theorem
CUAN	consistent uniformly asymptotically normal
d.f.	degrees of freedom
DGP	data-generating process
DW	Durbin–Watson

EM	estimation maximization (algorithm)
FCLT	functional CLT (invariance principle)
FGLS	feasible generalized least squares
FIML	full-information maximum likelihood
f.m.g.f.	factorial moment-generating function
GLS	generalized least squares
GMM	generalized method of moments
HCSE	heteroskedastic-consistent standard errors
I(d)	(fractionally) integrated process of order d
i.i.d.	independent and identically distributed
IID$_m(\boldsymbol{\theta})$, IID($\boldsymbol{\theta}$)	m-dimensional i.i.d. with parameter vector $\boldsymbol{\theta}$
ILS	indirect least squares
IMSE	integrated mean squared error
IV	instrumental variable
KLIC	Kullback–Leibler information criterion
LAD	least absolute deviations
LIL	law of iterated logarithm
LIML	limited-information maximum likelihood
LLN	law of large numbers; see also WLLN, SLLN
LM	Lagrange multiplier
LR	likelihood ratio
LS[E]	least squares [estimator]; see also 2SLS, 3SLS, CLS, FGLS, GLS, ILS, NLS, OLS, RLS, WLS
MA(q)	moving-average process of order q
[MC]MC	[Markov chain] Monte Carlo
m.g.f.	moment-generating function
ML[E]	maximum likelihood [estimator]; see also FIML, LIML, QML
MSE	mean squared error; see also IMSE
NLS	nonlinear least squares
OLS	ordinary least squares
p.d.f.	probability density function
QML[E]	quasi-maximum likelihood [estimator]
RLS	recursive least squares (*not* restricted LS)
RSS	residual sum of squares
r.v.	random variable
s.d.	standard deviation
s.e.	standard error
SLLN	strong law of large numbers
SUR[E]	seemingly unrelated regression [estimator]
UMP	uniformly most powerful
VAR	vector autoregressive, vector autoregression
W	Wald
WLLN	weak law of large numbers
WLS	weighted least squares

9. HOPES, FEARS AND EXPECTATIONS

Our hope is that this paper may contribute towards the establishment of a common notation in econometrics. Our fear is that it will not. We realize that it will be difficult to get consensus. The $=$ sign for equality was first proposed in the middle of the 16th century, but 150 years later Bernoulli still used $\propto$ (stylized æ, short for *aequalis*) in his *Ars Conjectandi*. Thus, our expectation is that it could take 150 years before a common notation is adopted.

ACKNOWLEDGEMENTS

We are grateful to the editors Pravin Trivedi and Neil Shephard and three referees of *The Econometrics Journal* for their positive and constructive comments on an unusual paper, and to our colleagues Robert Bartels, Peter Boswijk, Giuseppe Cavaliere, Russell Davidson, Walter Distaso, Jim Durbin, Arthur Goldberger, Clive Granger, David Hendry, Steve Lawford, Paolo Paruolo, and Peter Phillips for helpful comments. The first author acknowledges support from the ESRC (UK) grant R000239538.

REFERENCES

Beccari, C. (1997). Typesetting mathematics for science and technology according to ISO 31/XI. *TUGboat 18 no. 1*, 39–48.

Bollobás, B. (ed.) (1986). *Littlewood's Miscellany*. Cambridge: Cambridge University Press.

Kline, M. (1972). *Mathematical Thought From Ancient to Modern Times*. New York: Oxford University Press.

Magnus, J. R. and M. S. Morgan (1987). The ET interview: Professor J. Tinbergen. *Econometric Theory 3*, 117–42.

Mathematical Signs and Symbols for Use in the Physical Sciences and Technology, ISO 31/XI (1982). In *ISO Standards Handbook 2*. pp. 186–214. Geneva: International Organization for Standardization.

Symbols, Units and Nomenclature in Physics (1988). In R. C. West (ed.), *CRC Handbook of Chemistry and Physics*, F156–F193, 1st Student Edition, Boca Raton, Florida: CRC Press, Inc.

APPENDIX: FORWARD AND BACKWARD DIFFERENCING

The notation used in the 'difference calculus' branch of mathematics is $\triangle$ for forward differencing, and $\triangledown$ for backward differencing. For example, the derivative of a function $f(x)$ is defined as

$$\lim_{\triangle_h x \to 0} \frac{\triangle_h f(x)}{\triangle_h x} \equiv \lim_{h \to 0} \frac{f(x+h) - f(x)}{h},$$

whereas the random walk $y_t = y_{t-1} + \varepsilon_t$ is written as $\bigtriangledown y_t = \varepsilon_t$, where we assume that $\{\varepsilon_t\}_1^T \sim \text{IN}\left(0, \sigma^2\right)$ for expository purposes. The distinction between the two types of differencing is important because

$$\sum_{t=1}^{T} \frac{y_{t-1}}{\sigma\sqrt{T}} \cdot \frac{\varepsilon_t}{\sigma\sqrt{T}} = \sum_{t=1}^{T} \left(\frac{y_{t-1}}{\sigma\sqrt{T}}\right) \triangle \left(\frac{y_{t-1}}{\sigma\sqrt{T}}\right)$$
$$\xrightarrow{w} \int_0^1 W(s)\,\mathrm{d}W(s) = \frac{W(1)^2 - 1}{2}, \tag{A.1}$$

where $W(s)$ is the standard Wiener process on $s \in [0, 1]$. Notice that the definition of the differential requires the *forward* (not the backward) difference $\triangle y_{t-1} = \bigtriangledown y_t \neq \triangle y_t$, so one must be very careful in distinguishing forward from backward difference, and not using $\triangle$ for both within the same paper. Lest the reader think that it makes no difference, the convergence of the sum in (A.1) is conditional, and any evaluation point of $\triangle y_{\cdot}$ other than $t-1$ leads to a different integral. Letting $\bar{s}_j := a s_j + (1-a)\, s_{j-1}$ with $a \in [0, 1]$, and $s_\infty := s$,

$$\lim_{n\to\infty} \sum_{j=1}^{n} W(\bar{s}_j)(W(s_j) - W(s_{j-1}))$$
$$= \frac{1}{2}(W(s)^2 - W(s_0)^2) + \left(a - \frac{1}{2}\right)(s - s_0)$$

by stochastic integration. Clearly, the results depend on a. The case $a = 0$ gives the usual Itô integral of (A.1) whose value is $(W(1)^2 - 1)/2$, but $a = 1/2$ leads to the Stratonovich integral whose value is $W(1)^2/2$. Our suggested notation addresses the econometricians' reluctance to use $\bigtriangledown$, while keeping a clear distinction between forward and backward differencing.