Notation in Econometrics: A Proposal for a Standard

Karim M. Abadir
Department of Mathematics and Department of Economics, University of York, York, UK
E-mail: kma4@york.ac.uk

Jan R. Magnus
CentER, Tilburg University, Tilburg, The Netherlands
E-mail: magnus@uvt.nl

The Econometrics Journal, 5 (2002), 76–90.

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 and also obeys ISO regulations. 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.

Key words: Notation, Symbols, Econometrics, International Organization for Standardization (ISO).

JEL Code: C10

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 J.E. 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, many serious researchers *did* worry about notation. Jan Tinbergen propagated 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, with one major exception: the treatment of lowercase single-letter constants (such as the base of natural logarithms e and the imaginary unit i — very often written as e and i, contrary to ISO regulations) or operators (such as the derivative operator d — often written as d).[1] 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.

Notation matters. A good and consistent notation helps in the understanding, communication and development of our profession. 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

$$5\mathrm{p} : \mathrm{Rm} : 15$$
$$5\mathrm{m} : \mathrm{Rm} : 15$$
$$25\mathrm{m} : \mathrm{m} : 15 \text{ qd est } 40,$$

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

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, which is as close as possible to existing common practice and also obeys ISO regulations. 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.

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 LaTeX definitions are provided, and the complete list of definitions can be downloaded from *http://cwis.kub.nl/~few5/center/staff/magnus/notation.htp*.

2 Vectors and matrices

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

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

We write

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

for an $n \times 1$ vector $\boldsymbol{a}$ and an $m \times n$ matrix $\boldsymbol{A}$. If one has a choice, we recommend that $m \geq n$.

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 $\boldsymbol{a}_{\bullet 1}$, $\boldsymbol{a}_{\bullet 2}$, ..., $\boldsymbol{a}_{\bullet n}$, 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. The symbol $\bullet$ is produced by \bcdot since \cdot ($\cdot$) is too small and \bullet ($\bullet$) is too large. 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:

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

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 dimension of $\boldsymbol{x}$ equals the number of columns of $\boldsymbol{A}$ and the dimension of $\boldsymbol{b}$ equals the number of its rows. If this is the case then $\boldsymbol{A}$, $\boldsymbol{x}$ and $\boldsymbol{b}$ are conformable.

Two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ for which $\boldsymbol{a}'\boldsymbol{b} = 0$ are *orthogonal*. We also write $\boldsymbol{a} \perp \boldsymbol{b}$ (\bot). The column space of $\boldsymbol{A}$ is denoted $\mathrm{col}(\boldsymbol{A})$ (\col) and denotes the set $\{\boldsymbol{x} : \boldsymbol{x} = \boldsymbol{A}\boldsymbol{c} \text{ for some } \boldsymbol{c} \neq \mathbf{0}\}$. The null space of $\boldsymbol{A}$ is the set $\{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \mathbf{0}\}$. The null space of $\boldsymbol{A}'$ is denoted $\mathrm{col}^{\perp}(\boldsymbol{A})$ and is called the *orthogonal complement* of $\mathrm{col}(\boldsymbol{A})$. It defines the set $\{\boldsymbol{x} : \boldsymbol{A}'\boldsymbol{x} = \mathbf{0}\}$, which can also be written as $\{\boldsymbol{x} : \boldsymbol{x} \perp \boldsymbol{A}\}$.

3 Operations on matrix A and vector a

The following standard operations are proposed.

$\boldsymbol{A}'$	transpose
$\boldsymbol{A}^{-1}$	inverse
$\boldsymbol{A}^{+}$	Moore-Penrose inverse
$\boldsymbol{A}^{-}$	generalized inverse (satisfying $\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{A} = \boldsymbol{A}$)
$\mathrm{dg}\,\boldsymbol{A}$, $\mathrm{dg}(\boldsymbol{A})$	diagonal matrix containing the diagonal elements

	of $\boldsymbol{A}$ (\dg)
$\operatorname{diag}(a_1, \ldots, a_n)$	diagonal matrix containing $a_1, \ldots, a_n$
	on the diagonal (\diag)
$\operatorname{diag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n)$	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
$\boldsymbol{A}^p$	p-th power
$\boldsymbol{A}^{\#}$	adjoint (matrix)
$\boldsymbol{A}^*$	conjugate transpose
	(If $\boldsymbol{A} := \boldsymbol{U} + \mathrm{i}\boldsymbol{V}$ then $\boldsymbol{A}^* = \boldsymbol{U}' - \mathrm{i}\boldsymbol{V}'$)
$\boldsymbol{A}_k$	principal submatrix of order $k \times k$
$(\boldsymbol{A}, \boldsymbol{B})$, $(\boldsymbol{A} : \boldsymbol{B})$	partitioned matrix
$\operatorname{vec} \boldsymbol{A}$, $\operatorname{vec}(\boldsymbol{A})$	vec operator (\vec)
$\operatorname{vech} \boldsymbol{A}$, $\operatorname{vech}(\boldsymbol{A})$	vector containing a_{ij} $(i \geq j)$ (\vech)
$\operatorname{rk}(\boldsymbol{A})$	rank (\rk)
λ_i, $\lambda_i(\boldsymbol{A})$	i-th eigenvalue (of $\boldsymbol{A}$)
$\operatorname{tr} \boldsymbol{A}$, $\operatorname{tr}(\boldsymbol{A})$	trace (\tr)
$\operatorname{etr} \boldsymbol{A}$, $\operatorname{etr}(\boldsymbol{A})$	$\exp(\operatorname{tr} \boldsymbol{A})$ (\etr)
$\lvert\boldsymbol{A}\rvert$, $\det \boldsymbol{A}$, $\det(\boldsymbol{A})$	determinant (\det)
$\lVert\boldsymbol{A}\rVert$	norm of matrix $(\sqrt{(\operatorname{tr} \boldsymbol{A}^*\boldsymbol{A})})$ (\\|)
$\lVert\boldsymbol{a}\rVert$	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, \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)
$\boldsymbol{K}_{mn}$	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 $k \times k$

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.

If we have a symmetric matrix $\boldsymbol{A}$ of order $n \times n$, then the 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 where it is desirable that λ_1 denotes the largest eigenvalue.

4 The linear regression model

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

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

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, and in formulae involving the imaginary unit i confusion can be avoided by not using i as an index.

This formulation is not without controversy. Some authors write X_{ht} instead of x_{ih}, which is unsatisfactory, since $\boldsymbol{X}$ is an $n \times k$ matrix and hence in their formulation X_{ht} is the th-th 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.[2] If the errors are not spherical, we denote them by $\boldsymbol{u}$.

Estimators are random variables which say something about a fixed but unknown quantity, called a parameter. They are denoted by 'hats', such as $\hat{\boldsymbol{\beta}}$. (`\hat{\vbeta}`). If we have a second estimator of $\boldsymbol{\beta}$ this is denoted by a 'tilde': $\tilde{\boldsymbol{\beta}}$. 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 the 'prediction'.

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.[3] 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).

If $\boldsymbol{a}$ is a vector, say of order n, then $\bar{a}$ (`\bar`) denotes the average of its components: $\bar{a} = \boldsymbol{\imath}'\boldsymbol{a}/\boldsymbol{\imath}'\boldsymbol{\imath}$.

[2]The vector $\boldsymbol{\varepsilon}$ is spherically distributed if $\boldsymbol{\varepsilon}$ and $\boldsymbol{H}\boldsymbol{\varepsilon}$ have identical distributions for every orthogonal matrix $\boldsymbol{H}$.

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

It is customary to write

$$\boldsymbol{P_X} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{+}\boldsymbol{X}', \qquad \boldsymbol{M_X} = \boldsymbol{I}_n - \boldsymbol{P_X}$$

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

$$\boldsymbol{M_\imath} = \boldsymbol{I}_n - (1/n)\boldsymbol{\imath}\boldsymbol{\imath}',$$

and the vector $\boldsymbol{M_\imath a}$ denotes the vector $\boldsymbol{a}$ in deviation from its mean.

We denote a null hypothesis as H_0 (\rH) and an alternative as H_A (not H_a since a may be a scalar or may refer to 'asymptotic'). The statement of H_0: $\boldsymbol{R}'\boldsymbol{\beta} = \boldsymbol{c}$ is preferred over $\boldsymbol{R\beta} = \boldsymbol{r}$. 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 (dimension of $\boldsymbol{c}$). 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

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

where $\boldsymbol{\Omega} = \sigma^2\boldsymbol{V}$. We prefer the use of $\boldsymbol{\Omega}$ over $\boldsymbol{\Sigma}$, which can be confused with the summation symbol. If the variance matrix $\boldsymbol{\Omega}$ contains parameters, we write $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\alpha})$. The dimension 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 dimension m.

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

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

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

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

where $\boldsymbol{y}_{i\boldsymbol{\cdot}}$ and $\boldsymbol{u}_{i\boldsymbol{\cdot}}$ 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:

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

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

5 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.

6 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 above in some neighborhood of c (possibly $\pm\infty$), 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'.[4] 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 p.d.f. and c.d.f. of the standard-normal distribution, 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 (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 by x in $a \leq x < b$ are denoted by $[a, b)$. Occasionally it might be unclear whether (a, b) indicates a real-line interval or a

[4] 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'.

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, 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 $\boldsymbol{Z}_m$, $\boldsymbol{Z}_{m+1}$,...,$\boldsymbol{Z}_n$.

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
$B \backslash A$	$B \cap A^c$
$\mathring{S}$	interior of S (`\interior{S}`)
S'	derived set of S
$\bar{S}$	closure of S (`\bar{S}`)
∂S	boundary of S (`\partial S`)

We denote functions by

$f : S \to T$	function defined on S with values in T
f, g, φ, ψ, ϑ	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 $(g * f)(x) = \int_{-\infty}^{\infty} g(y) f(x - y)\, \mathrm{d}y$

For their differentials, derivatives and differences, we write

d	differential (`\rd`)
d^n	n-th order differential
$\mathrm{D}_j \varphi(\boldsymbol{x})$	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, $\partial \mathrm{D}_j f_i(\boldsymbol{x})/\partial x_k$
$\varphi^{(n)}(x)$	n-th 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 $\boldsymbol{F}(\boldsymbol{X})$
$\partial \operatorname{vec}\boldsymbol{F}(\boldsymbol{X})/\partial(\operatorname{vec}\boldsymbol{X})'$	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	(backward) difference operator: $\Delta x_t = x_t - x_{t-1}$ (`\diff`)
$\triangle$	forward difference operator: $\triangle x_t = x_{t+1} - x_t$ (`\fordiff`)
$[f(x)]_a^b$, $f(x)\|_a^b$	$f(b) - f(a)$

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. Notice the difference between the differencing operator `\diff` (Δ) and the gradient `\nabla` (∇).

We use L (or B) rather than $\mathcal{L}$ for the lag operator in order to avoid confusion with the Laplace transform. This and other useful transforms are defined by

$\mathcal{F}\{\cdot\}$	Fourier transform (`\calF`)
$\mathcal{F}^{-1}\{\cdot\}$	inverse Fourier transform
$\mathcal{L}\{\cdot\}$	Laplace transform (`\calL`)
$\mathcal{L}^{-1}\{\cdot\}$	inverse Laplace transform
$\mathcal{M}\{\cdot\}$	Mellin transform (`\calM`)
$\mathcal{M}^{-1}\{\cdot\}$	inverse Mellin 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`)
$\lvert x \rvert$	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
$B(\boldsymbol{c})$, $B(\boldsymbol{c}; r)$, $B(\boldsymbol{C}; r)$	neighborhood (ball) with center $\boldsymbol{c}$ ($\boldsymbol{C}$) 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 $\mathbb{R}^{n \times k}$.

7 Statistical symbols, functions and operators

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`)
$\mathrm{E}(\cdot\lvert\cdot)$	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)
$\boldsymbol{\mathcal{I}}$	(Fisher) information matrix (\Infmat)
$\mathcal{F}_t$	filtration at time t (\calF)
t	t-statistic, t-value
$\to$, $\longrightarrow$	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. The matrix $-\boldsymbol{\mathcal{H}}$ is also called the observed information matrix, while its expectation $\boldsymbol{\mathcal{I}} := -\,\mathrm{E}(\boldsymbol{\mathcal{H}})$ is the expected information matrix.

The main distributions in statistics are denoted as follows.

$\mathrm{Bin}(n,p)$	binomial distribution (\Bin)
$\mathrm{Poi}(\mu)$	Poisson distribution (\Poi)
$\mathrm{U}(a,b)$	uniform distribution (\rU)
$\mathrm{N}_m(\boldsymbol{\mu},\boldsymbol{\Omega})$	m-dimensional normal distribution (\rN)
$\mathrm{LN}(\mu,\sigma^2)$	lognormal distribution (\LN)
$\phi(\cdot)$	standard-normal p.d.f. (\phi)
$\Phi(\cdot)$	standard-normal c.d.f.
$\mathrm{IN}_m(\boldsymbol{\mu}_i,\boldsymbol{\Omega}_i)$	sequence $i=1,2,\ldots$ of independent m-dimensional normal distributions
$\chi^2_n(\delta)$	chi-squared distribution with n d.f. and non-centrality parameter δ.
χ^2_n	central chi-squared ($\delta=0$)

$\mathrm{t}_n(\delta)$	Student distribution with n d.f. and noncentrality δ (\rt)
t_n	central t ($\delta = 0$)
$\mathrm{C}(a, b)$	Cauchy distribution (\rC)
$\mathrm{F}_{m,n}(\delta)$	Fisher distribution with m (numerator) and n (denominator) d.f. and non-centrality δ (\rF)
$\mathrm{F}_{m,n}$	central F ($\delta = 0$)
$\Gamma(\alpha, \lambda)$	gamma distribution
$\mathrm{B}(a, b)$	beta distribution (\rB)
$W(\tau)$, $B(\tau)$	standard Wiener process (Brownian motion) on $\tau \in [0, 1]$

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 could indicate either is best avoided. Like 'expectation', the words 'variance' (var), 'covariance' (cov), and 'correlation' (corr) indicate population parameters. The corresponding sample parameters are called 'sample variance', 'sample covariance' and 'sample correlation'.

The 'standard deviation' is the positive square root of the variance. If θ is a parameter which we estimate by $\hat{\theta}$, then this estimator is a random variable with a variance $\mathrm{var}(\hat{\theta})$ and a standard deviation $\sqrt{\mathrm{var}(\hat{\theta})}$. In general, this standard deviation depends on unknown parameters. Both the estimator of the standard deviation and its realization are called the 'standard error'. The t-statistic is a random variable (not necessarily Student distributed); its realization is the t-value.

8 Abbreviations and acronyms

2SLS	two-stage least squares
3SLS	three-stage least squares
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
c.d.f.	cumulative distribution function
c.f.	characteristic function
c.g.f.	cumulant-generating function
CLT	central limit theorem
CUAN	consistent uniformly asymptotically normal
d.f.	degrees of freedom
DW	Durbin-Watson
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
i.i.d.	independent and identically distributed
ILS	indirect least squares
I(d)	(fractionally) integrated process of order d
IV	instrumental variable
LAD	least absolute deviations
LIL	law of iterated logarithm
LIML	limited-information maximum likelihood
LLN	law of large numbers
LM	Lagrange multiplier
LR	likelihood ratio
LS[E]	least squares [estimator]; see also 2SLS, 3SLS, FGLS, GLS, ILS, NLS, OLS, RLS
MA(q)	moving-average process of order q
m.g.f.	moment-generating function
ML[E]	maximum likelihood [estimator]; see also FIML, LIML, QML
MSE	mean squared error
NLS	nonlinear least squares
OLS	ordinary least squares

p.d.f.	probability density function
QML[E]	quasi-maximum likelihood [estimator]
RLS	restricted least squares
r.v.	random variable
s.e.	standard error
SUR	seemingly unrelated regression
UMP	uniformly most powerful
W	Wald

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 Paolo Paruolo for helpful comments.

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. Tin-

bergen, *Econometric Theory 3*, 117–142.

Mathematical Signs and Symbols for Use in the Physical Sciences and Technology, ISO 31/XI (1982). In *ISO Standards Handbook 2*, 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.