A matrix extension of the Cauchy-Schwarz inequality

Gautam Tripathi

Department of Economics, University of Wisconsin, Madison WI-53706, USA

Received 1 September 1998; accepted 4 December 1998

Abstract

A simple argument is used to obtain a very useful generalization of the well known Cauchy-Schwarz inequality. © 1999 Elsevier Science S.A. All rights reserved.

Keywords: Cauchy-Schwarz Inequality

JEL classification: C10

1. Main result

Notation 1. Let A and B be two \( p \times p \) matrices. We write \( A \preceq B \) if and only if \( B - A \) is non-negative definite. \( \| A \| \) denotes the Euclidean norm of a matrix; i.e. \( \| A \| = \sqrt{\sum_{i,j=1}^{p} a_{i,j}^2} \). The transpose of A is given by \( A' \). □

Using this notation we prove the following, very useful, inequality.

Theorem 1. Let \( x \in \mathbb{R}^p \) and \( y \in \mathbb{R}^q \) be random vectors such that \( E\| x \|^2 < \infty \), \( E\| y \|^2 < \infty \), and \( Eyy' \) is non-singular. Then,

\[
(Exy'(Ey)' - 1)(Exx') \preceq Exx'.
\]

Example 1.1. (A Simple Application). As an application of the above inequality consider the regression model \( y = \theta'x + \gamma'z + \epsilon \) with i.i.d. observations. Here \( \epsilon \) is independent of \( (x,z) \), has mean zero and variance \( \sigma^2 \). Subtracting \( E(y|z) \) from both sizes of this regression model we obtain the ‘difference’ model \( y - E(y|z) = \theta'[x - E(x|z)] + \epsilon \). Therefore, doing nonparametric least squares on this differenced model (Robinson, 1988), we get that \( n^{1/2}(\hat{\theta}_{NPLS} - \theta) \overset{d}{\to} N(0,\sigma^2\Phi^{-1}) \), where \( \Phi = Exx' - E[E(x|z)E(x'|z)] \). Furthermore, as is well known, \( n^{1/2}(\hat{\theta}_{LS} - \theta) \overset{d}{\to} N(0,\sigma^2(I_{xx} - I_{xz}I_{zz}^{-1}I_{xz}))^{-1} \) where \( \hat{\theta}_{LS} \) is the least squares estimator of \( \theta \) (in the presence of \( \gamma \)), \( I_{xx} = Exx' \), \( I_{xz} = Exz' \), and

*Tel.: +608-262-3804.
E-mail address: gtripath@ssc.wisc.edu (G. Tripathi)

0165-1765/99/$ – see front matter © 1999 Elsevier Science S.A. All rights reserved.
PII: S0165-1765(99)00014-2
\[ I_\alpha = \mathbb{E}zz'. \] As estimating \( \theta \) in the differenced model implies that we are discarding information provided by the linear index intuition suggests that \( \hat{\theta}_{NLS} \) should be at least as efficient as \( \hat{\theta}_{NPLS} \). That this is indeed so can be seen immediately by writing \( \mathbb{E}(xz') = \mathbb{E}(\mathbb{E}(x|z)z') \), and using Theorem 1.1. Robinson (1988, Page 941) makes a similar point when he refers to the “Schwarz” inequality, but he does not elaborate. This note, therefore, also clarifies Robinson’s reference to a Schwarz inequality. \( \square \)

**Proof of Theorem 1.1.** Choose any \((\alpha, \delta) \in (\mathbb{R}^p \times \mathbb{R}^q)\). Now let \( i(\alpha, \delta) = \mathbb{E}(x'\alpha + y'\delta)^2 \), and \( \delta^* = \arg \min_{\delta \in \mathbb{R}^q} i(\alpha, \delta) \); i.e. \(-y'\delta^*\) is the projection of \(x'\alpha\) onto \(\text{span}\{y_1, \ldots, y_p\}\). From Lemma 1.1. we know that \(\delta^* = - (\mathbb{E}yy')^{-1}(\mathbb{E}yx')\alpha\). Hence, \(x'\alpha + y'\delta^* = [x' - y'(\mathbb{E}yy')^{-1}(\mathbb{E}yx')]\alpha\). Therefore, keeping in mind that \(x'\alpha + y'\delta^*\) is a scalar,

\[
i(\alpha, \delta^*) = \mathbb{E}(x'\alpha + y'\delta^*)^2 = \mathbb{E}(x'\alpha + y'\delta^*)(x'\alpha + y'\delta^*)
\]

\[
= \alpha'\mathbb{E}[x - (\mathbb{E}yx') \mathbb{E}yy')^{-1}y][x' - y'(\mathbb{E}yy')^{-1}(\mathbb{E}yx')]\alpha
\]

\[
= \alpha'\{\mathbb{E}xx' - (\mathbb{E}yx')(\mathbb{E}yy')^{-1}(\mathbb{E}yx')\} \alpha.
\]

The result follows since \(i(\alpha, \delta^*) \geq 0\), and \(\alpha\) was chosen arbitrarily. \( \square \)

**Remark 1.1.** The inequality is sharp if \(x'a + y'b = 0\) w.p. 1, for some \((a, b) \in \mathbb{R}^p \times \mathbb{R}^q\); i.e. if the distribution of \((x, y)\) puts all its mass in some proper linear subspace of \(\mathbb{R}^{p+q}\). Also notice that when \(p = q = 1\), i.e. when we are dealing with random variables, the above result reduces to the usual Cauchy-Schwarz inequality; i.e. \((\mathbb{E}xy)^2 \leq \mathbb{E}x^2\mathbb{E}y^2\).

**Lemma 1.1.** Define \(\delta^*\) as in Theorem 1.1; i.e. \(\delta^* = \arg \min_{\delta \in \mathbb{R}^q} \mathbb{E}(x'\alpha + y'\delta)^2\). Then, \(\delta^* = - (\mathbb{E}yy')^{-1}(\mathbb{E}yx')\alpha\).

**Proof.** Clearly \(\delta^* \in \mathbb{R}^q\). So all that remains is to verify the orthogonality condition of the classical projection theorem. To see that the orthogonality condition holds, pick any \(\delta \in \mathbb{R}^q\). Then, keeping in mind that \(y'(\mathbb{E}yy')^{-1}(\mathbb{E}yx')\alpha\) is a scalar,

\[
\mathbb{E}\{x'\alpha + y'\delta^*\}(y'\delta)
\]

\[
= \mathbb{E}\{x'\alpha - y'(\mathbb{E}yy')^{-1}(\mathbb{E}yx')\alpha\}(y'\delta)
\]

\[
= \mathbb{E}\{\alpha'x - \alpha'(\mathbb{E}yx')(\mathbb{E}yy')^{-1}y\}(y'\delta)
\]

\[
= \mathbb{E}\{(\alpha'x)(y'\delta)\} - \alpha'(\mathbb{E}yx')\mathbb{E}yy')^{-1}(\mathbb{E}yx')\delta
\]

\[
= \alpha'(\mathbb{E}yx')\delta - \alpha'(\mathbb{E}yx')\delta = 0.
\]

Hence, we are done. \( \square \)

2. **Conclusion**

This inequality was obtained in response to a question asked by some students while I was teaching a graduate course in Econometrics. I have found it to be a handy little tool while looking at issues
related to asymptotic efficiency of estimators. Although this inequality looks astonishingly familiar, I have been unable to discover any references to it in the literature that I have reviewed.

References