



Influence of the bin-to-bin and inside bin correlations on measured quantities

V. Perevoztchikov¹, M. Tabidze² and A. Tomaradze³

¹ Brookhaven National Laboratory, Upton, USA

² Institute for High Energy Physics of Tbilisi State University, Georgia

³ Northwestern University, Evanston, USA. e-mail Amiran@fnal.gov

Abstract

A new method for measuring the quantities influenced by bin-to-bin and inside bin correlations is presented. This method is essential for large multiplicity and/or high density of particles in phase space. The method was applied to the two particle correlation functions of $e^+e^- \rightarrow W^+W^-$ events.

1 Introduction

The usual calculations of statistical errors for entries in histograms and the use of these errors in the fitting procedure can bias the measurements if there are several entries from the same event. Traditionally, in the past this problem was ignored. The effect is small for low multiplicity events. However, for LEP and especially for the future LHC, RHIC, etc... experiments this effect is not small at all. These entries are correlated creating bin-to-bin and inside bin correlations. Neglecting these correlations, as we will show below, leads to a remarkable underestimation of the errors in the measured quantities and less precise estimation of the quantities themselves.

An effective approach has been proposed in [1]. The exact method described here includes all possible correlation effects.

The method was applied to the two-particle correlation functions of $e^+e^- \rightarrow W^+W^-$ events.

2 The Method

In this section a new method is presented for construction of the covariance matrix between the bins of the histogram for an unbiased measurement of the fitted quantities.

The presence of bin-to-bin correlations in two particle distributions is unavoidable. If there are N positive tracks in the events, each of them has $(N - 1)$ entries in the two-particle density P , contributing to different bins of the histogram. Also, due to the finite bin width, the same track can also enter several times in the same bin, which is a source of inside bin correlations.

The method is based on classical statistics. Let us consider the i -th event from the set of n events and the two-particle density P which is presented in the histogram h^i with N_p bins.

The histogram $H = \sum_{i=1}^n h^i$ and values

$$c_{jk} = \sum_{i=1}^n (h_j^i - H_j/n)(h_k^i - H_k/n)(1 + 1/n)$$

were calculated event by event. Here j and k are the bin numbers for the histograms. We do not know the correlations and errors for one event. But we know that the different events are uncorrelated. Let us consider bin values of the histogram made for one event as an random vector with unknown distribution. We have an uncorrelated ensemble of these vectors and hence we can estimate the covariance matrix statistically. It is important to note, that this algorithm computes both the "technical" and the physical correlations. For all events we have the resulting histogram H for the two-particle density P and $V_{jk} = c_{jk} \cdot n/(n - 1)$ covariance matrix for this histogram. Note, that the expression for unbiased estimation of sample covariance was used for the calculation of covariance between two values[2]. The diagonal terms V_{jj} of the covariance matrix are assumed to be the estimate for the squares of the error σ_j of the j -th of histogram H .

3 The Test of the Method Using the Simulated Events

The precise measurement of the correlations between particles became important for $e^+e^- \rightarrow W^+W^-$ events due to possible large impact of these correlations on the measured W mass [3]. Thus, we applied the method to the correlation functions in fully hadronic and semileptonic WW events.

The correlation function R is used to study the enhanced probability for emission of particles. For pairs of particles, it is defined as

$$R(p_1, p_2) = \frac{P(p_1, p_2)}{P_0(p_1, p_2)}, \quad (1)$$

where $P(p_1, p_2)$ is the two-particle probability density, p_i is the four-momentum of particle i , and $P_0(p_1, p_2)$ is a reference two-particle distribution which, ideally, resembles $P(p_1, p_2)$ in all respects, apart from the lack of Bose-Einstein symmetrization. The effect is usually described in terms of the variable Q , defined by $Q^2 = -(p_1 - p_2)^2 = M^2(\pi\pi) - 4m_\pi^2$, where M is the invariant mass of the two pions. The correlation function can then be written as

$$R(Q) = \frac{P(Q)}{P_0(Q)}, \quad (2)$$

which is frequently parametrised by the function

$$R(Q) = 1 + \lambda e^{-r^2 Q^2}. \quad (3)$$

In the above equation, in the hypothesis of a spherically symmetric pion source, the parameter r gives the RMS radius of the source and λ is the strength of the correlation between the pions.

The method described in section 2 was tested on the JETSET simulation [3, 4]. We choose the fully hadronic and semileptonic decay of WW pairs with Bose-Einstein correlations included using the LUBOEI code. The value of $\lambda = 0.85$ and $r = 0.5$ fm was used. In case of fully hadronic channel the Bose-Einstein correlations were switched on for all pions (full Bose-Einstein correlations).

The correlation matrices $\rho_{jk} = V_{jk}/(\sigma_j\sigma_k)$ for like-sign pairs of WW semileptonic channel (referred as $(2q)$ mode) and for fully hadronic channel (referred as $(4q)$ mode), computed using 100 000 simulated events for each sample, are shown in Fig. 1 and Fig. 2. The correlations between bins for the WW fully hadronic channel (Fig. 2) are larger than in the mixed hadronic and leptonic channel (Fig. 1). Thus, the bin-to-bin correlations are increased with multiplicity, as expected. Notice that the correlations are nearly independent of Q , which shows why the effective approach of [1] is a good approximation in this case.

For the future analysis the 500 samples of 3000 events each for $(4q)$ channel and 500 samples of 1500 events each for $(2q)$ channel were simulated. For each of these samples a histogram of the correlation function $R(Q)$ was built, using the 25 bins of 100 MeV from 0 to 2.5 GeV. The simulated R distributions were normalized to unity in the region $Q > 0.8$ GeV/ c^2 . We performed a χ^2 fit to the $R(Q)$ to the form (3) for each of 500 samples.

The average values of λ and r from these ‘naive’ fits were:

$$\lambda_{2q} = 0.333 \pm 0.029, \quad (4)$$

$$r_{2q} = 0.562 \pm 0.033 \text{ fm} \quad (5)$$

for $(2q)$ events and

$$\lambda_{4q} = 0.416 \pm 0.013, \quad (6)$$

$$r_{4q} = 0.561 \pm 0.012 \text{ fm} \quad (7)$$

for $(4q)$ events. The statistical errors correspond to the average of the errors for 500 samples. The “pull” of the fitted values of λ and r for the simulated $(4q)$ events are shown in Fig. 3. A Gaussian fits gave that the errors in parameters λ and r are underestimated by a factor 1.20 ± 0.05 and 1.30 ± 0.05 for the $(2q)$ events and by a factor 1.42 ± 0.06 and 1.53 ± 0.06 for the $(4q)$ events.

The average values of λ and r from the 500 fits using the inverted V_{jk} matrix (calculated for each of 500 samples) were:

$$\lambda_{2q} = 0.332 \pm 0.034, \quad (8)$$

$$r_{2q} = 0.556 \pm 0.040 \text{ fm} \quad (9)$$

for $(2q)$ events and

$$\lambda_{4q} = 0.403 \pm 0.017, \quad (10)$$

$$r_{4q} = 0.565 \pm 0.016 \text{ fm} \quad (11)$$

for $(4q)$ events. The “pull” of the fitted values of λ and r for the simulated $(4q)$ events are shown in Fig. 4. A Gaussian fits gave $\sigma_{(\lambda-\langle\lambda\rangle)}/\sigma_\lambda=1.06 \pm 0.05$ and $\sigma_{(r-\langle r\rangle)}/\sigma_r=1.02 \pm 0.05$ for $(2q)$ channel, and $\sigma_{(\lambda-\langle\lambda\rangle)}/\sigma_\lambda=0.96 \pm 0.05$ and $\sigma_{(r-\langle r\rangle)}/\sigma_r=1.08 \pm 0.05$ for $(4q)$ channel. The above values are in a good agreement with unity and thus the errors are correctly estimated.

4 Analysis of the Data

The data used in this section were collected with the DELPHI detector [5] at LEP2 in 1997, 1998 and 1999 at centre-of-mass energies from 183 to 204 GeV. The total integrated luminosities used in the present analyses was about 437 pb^{-1} . The number of selected WW events was 1452 for semileptonic channel and 2894 for fully hadronic channel.

The correlation matrices $\rho_{jk} = V_{jk}/(\sigma_j\sigma_k)$ for WW semileptonic channel and for the fully hadronic decay channel, computed from the data themselves, are shown in Fig. 5 and in Fig. 6.

The $R_{2q}(Q)$ and $R_{4q}(Q)$ distributions are shown in Fig. 7. The errors in the figures were obtained by the method described in Section 2. A fit to the correlation functions $R(Q)$ by the expression (3) using the inverted V_{jk} matrix yielded the following values:

$$\lambda_{2q} = 0.281 \pm 0.035, \quad (12)$$

$$r_{2q} = 0.589 \pm 0.047 \text{ fm} \quad (13)$$

for $(2q)$ events and

$$\lambda_{4q} = 0.286 \pm 0.018, \quad (14)$$

$$r_{4q} = 0.610 \pm 0.028 \text{ fm} \quad (15)$$

for $(4q)$ events. The fit results are shown by the curves in Fig. 7. Note that the values (12)–(15) were not corrected on detector effects, and, consequently they can not be directly compared to the values (8)–(11) obtained for the simulated events.

It has to be noted that the fit of the ratio of correlation functions of like-sign to unlike-sign particle pairs for data, as well for simulated events, yielded practically the same values and errors with and without using the covariance matrix in the fitting procedure. This statement is valid only if the correlation functions for both like-sign and unlike-sign pairs were obtained from the same data events. A similar effect was found in Ref. [6].

5 Summary

A model independent method for measuring the quantities influenced by bin-to-bin and inside bin correlations is described. A package, as an addition to HBOOK, was written to support this new functionality. The method was tested using the simulated WW events.

The values of correlation strength λ and radius r were measured for WW semileptonic and fully hadronic events with the proposed method using the data collected by the DELPHI detector at LEP2.

References

- [1] A. De Angelis and L. Vitale, Nucl. Instr. Methods **A423** (1999) 446.
- [2] A. Korn and M. Korn. "Mathematical handbook for scientist and engineers", New York, 1968.
- [3] L. Lönnblad and T. Sjöstrand, Eur. Phys. J. **C2** (1998) 165.
- [4] T. Sjöstrand, *PYTHIA 5.7 / JETSET 7.4*, CERN-TH.7112/93 (1993).
- [5] P. Abreu et al. (DELPHI Coll), Nucl. Instr. & Meth. **A378** (1996) 57.
- [6] R. Barate et al. (ALEPH Coll), CERN-EP/99-173(1999). Accepted by Phys. Lett. B.

PYTHIA $\lambda=0.85$

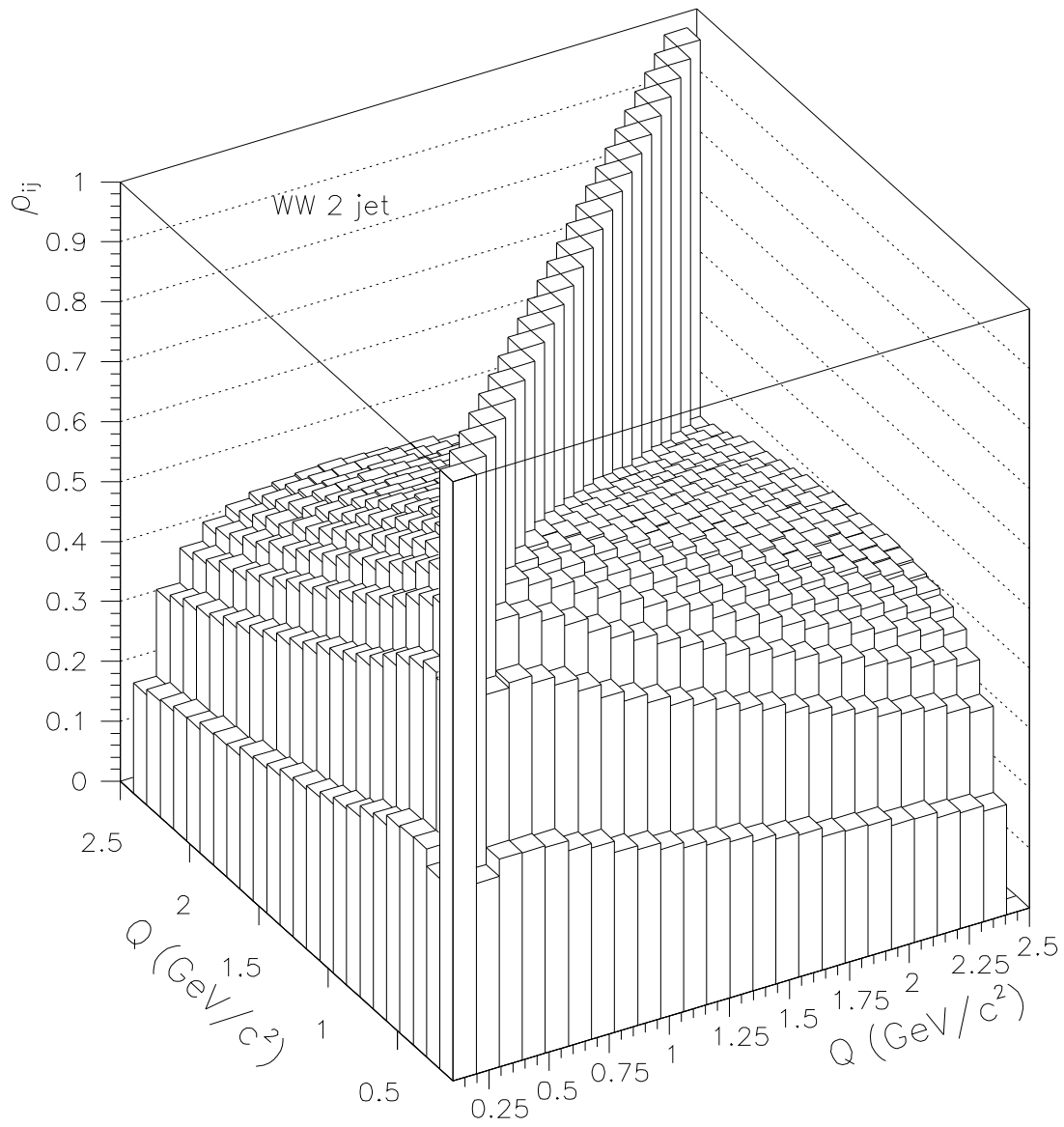


Figure 1: The correlation matrix for like-sign pairs obtained using the simulated WW ($2q$) events.

PYTHIA $\lambda=0.85$

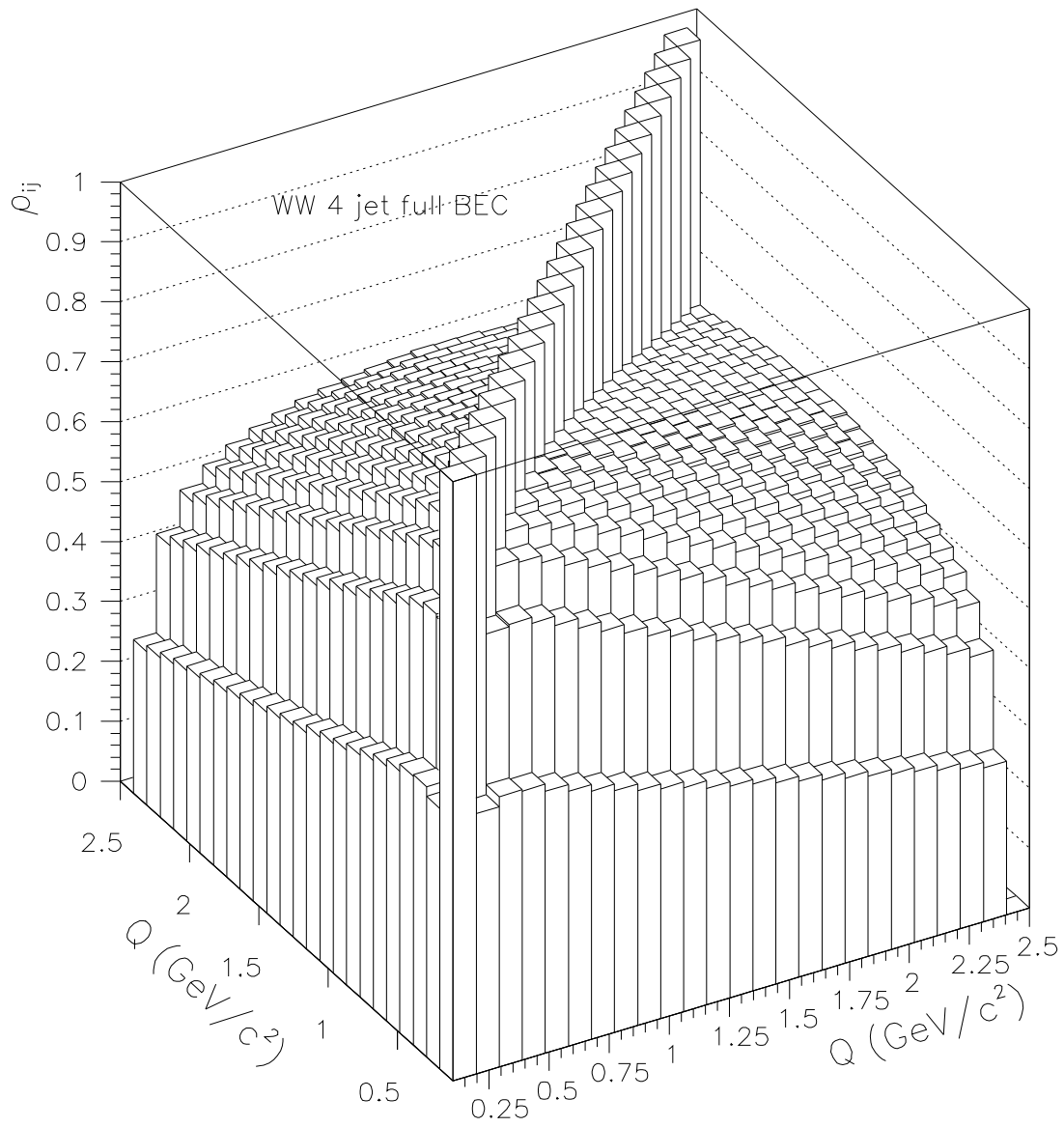


Figure 2: The correlation matrix for like-sign pairs obtained using the simulated WW ($4q$) events.

PYTHIA full BEC $\lambda=0.85$

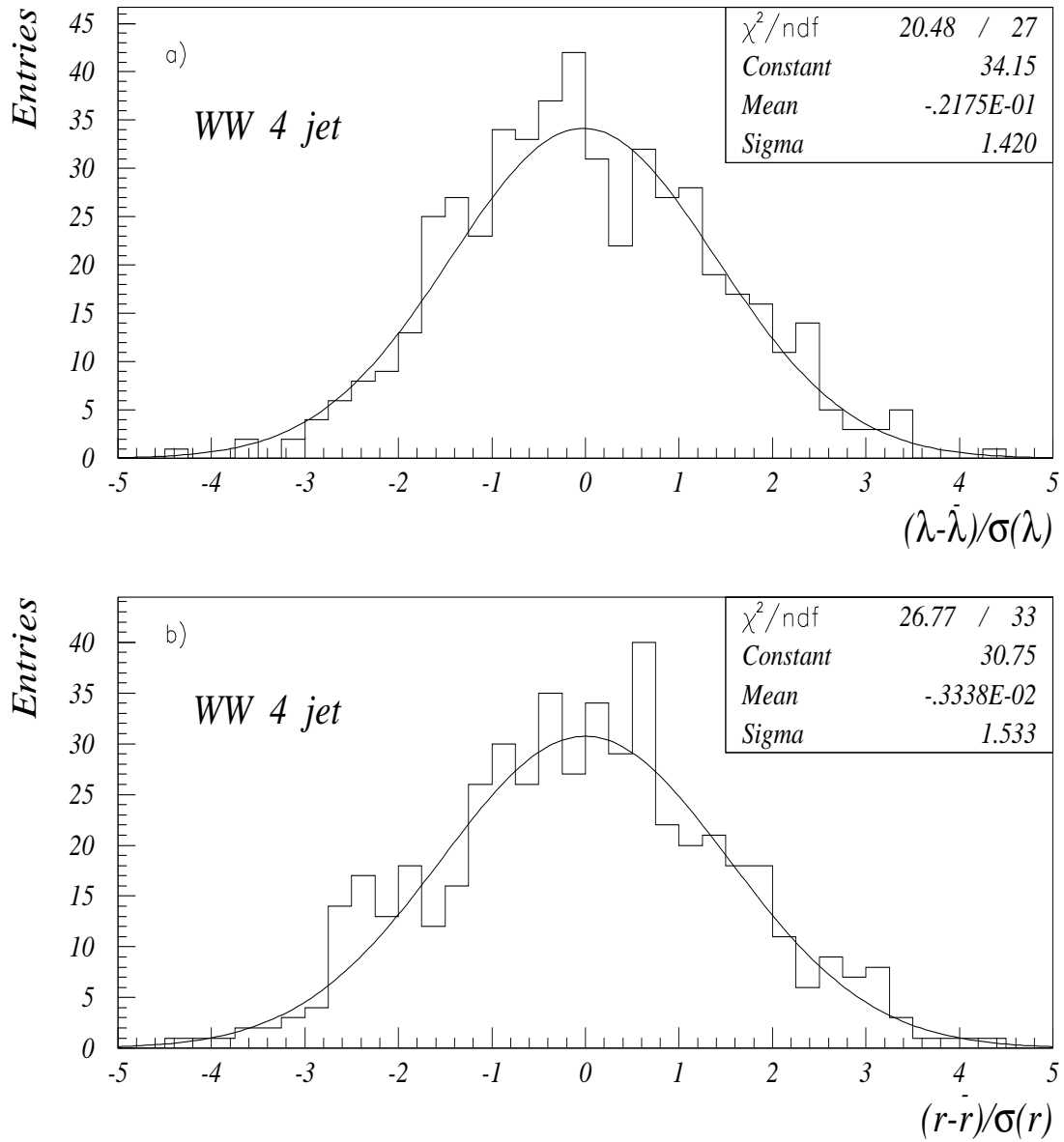


Figure 3: (a) Pull function for fitted parameter λ using a binned uncorelated least squares fit in the simulated samples. A gaussian fit is superimposed as a solid line. b) Same as (a) but for the parameter r .

PYTHIA full BEC $\lambda=0.85$

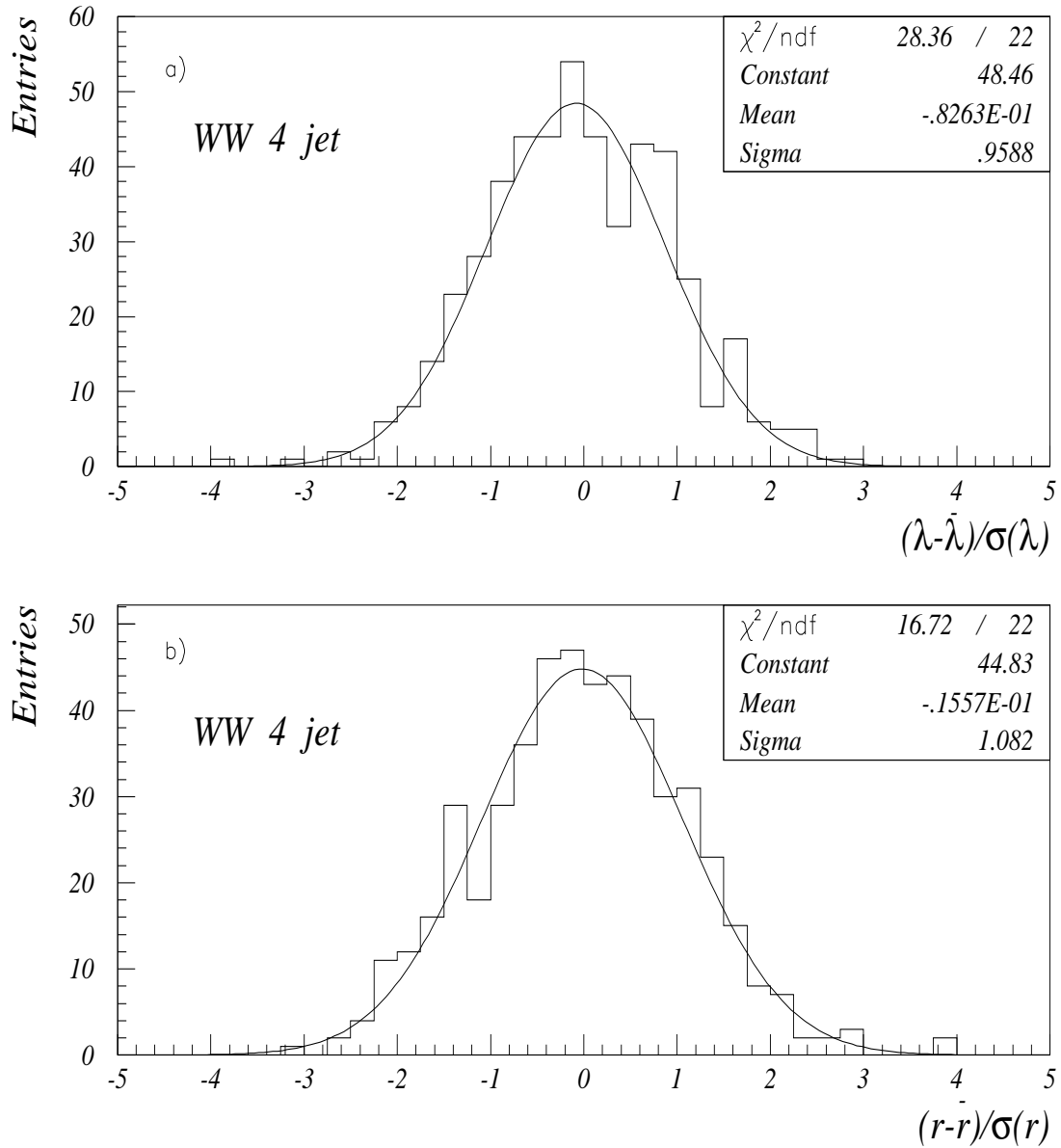


Figure 4: (a) Pull function for fitted parameter λ in the simulated samples using covariance matrix technique. A gaussian fit is superimposed as a solid line. b) Same as (a) but for the parameter r .

DELPHI

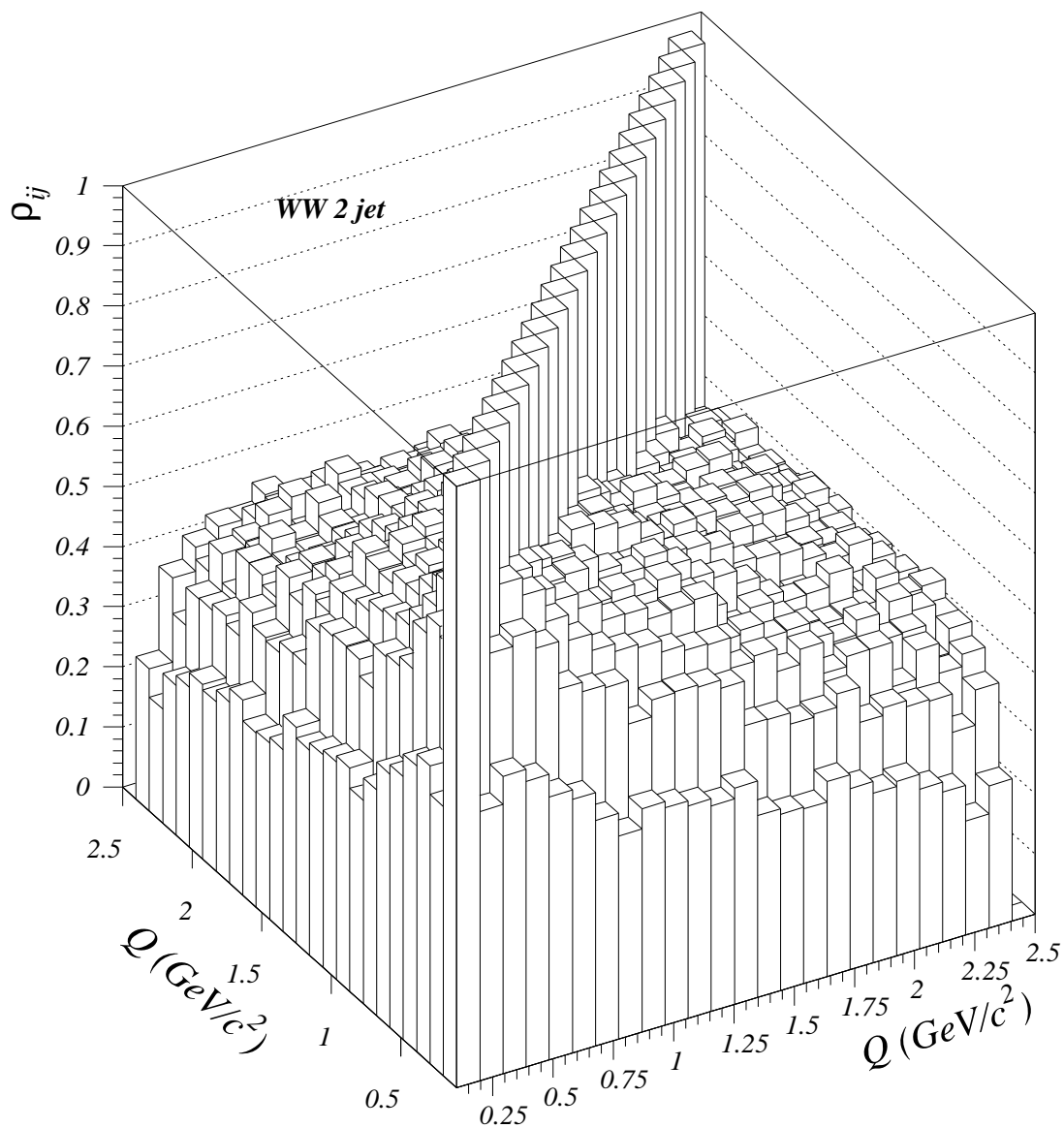


Figure 5: The measured correlation matrix for like-sign pairs of the WW ($2q$) channel.

DELPHI

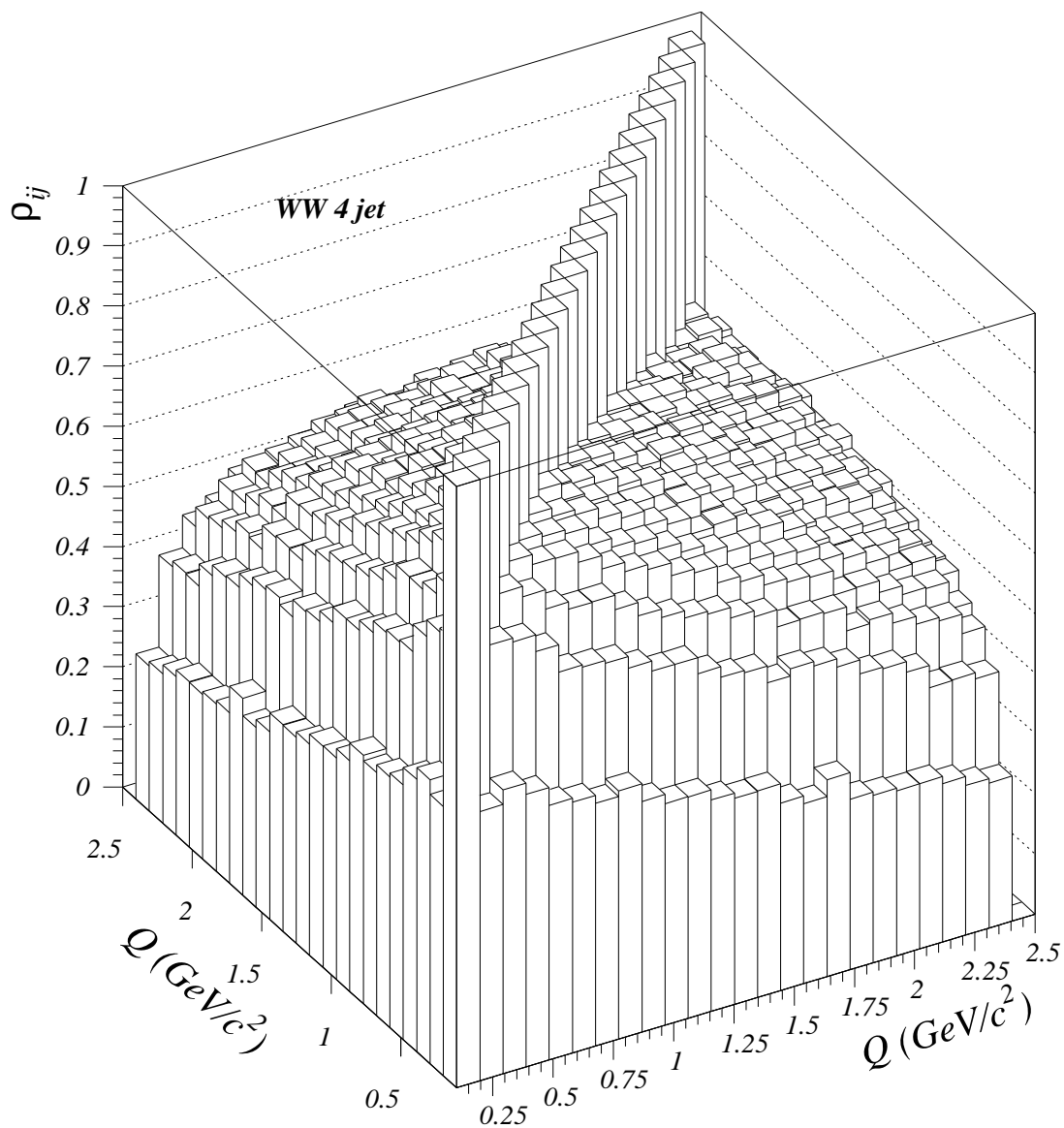


Figure 6: The measured correlation matrix for like-sign pairs of the WW (4q) channel.

DELPHI

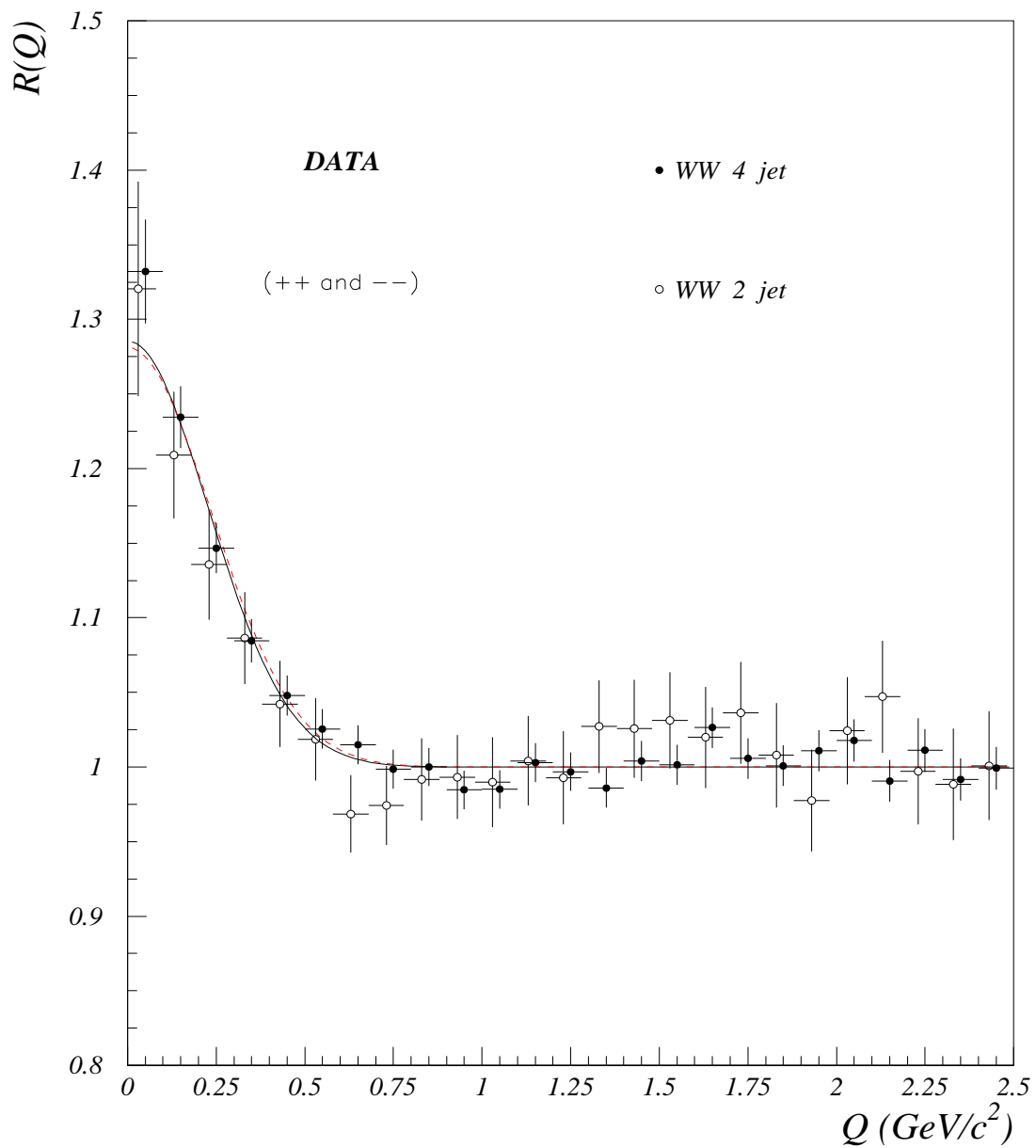


Figure 7: The measured correlation functions $R_{2q}(Q)$ (open circles) and $R_{4q}(Q)$ (closed circles) for like-sign pairs. The curves are the fit results using expression (3).