Theoretical Physics Lab-Course 2021S
University of Vienna

Unbiased Fitting

Milutin Popovic
Tim Vogel

Supervisor: Peter Stoffer

April 18, 2021

Abstract

When provided with data that doesn't only come with statistical uncertainties but also with systematic uncertainties, the least square method will not be sufficient enough and will provide false results. This report will emphasize on the D'Agostini bias, which explains the corrrelation between these uncertainties and give an example of data provided by particle physics measurements where the bias is present. In this regard the report will also explain how to avoid the D'Agostini bias with the t_0-method.

Contents

1 Introduction and Motivation

In physics, we often come across the situation, where measured data needs to be approximated by a theoretical model, which requires a set amount of parameters. To determine these parameters a so called "data fit" is required, which can be done via different techniques. One of the most widely used fit-techniques is "Least-squares fitting", but as soon as we deal with correlated data points, which means, each data point is not a completely independent measurement, a wrong application of the Least-squares fit, can lead to a bias, which, in return, will affect the accuracy of the fit in a negative way. This so called D'Agostini bias, albeit a very situational phenomenon, has to be considered when dealing with correlated data from one or even more experiments. It can be avoided by implementing a iterative fit method and we will consider this in an example from particle physics. The pion Vector Form Factor is a perfect example for this bias, as the experimental (even after many exact experiments) still doesn't fit the predictions of the theoretical model perfectly. In this report we will fit experimental data to the theoretical model, by implementing the so called "$t_0 - model$" into the fit.

2 Physical background and Findings

2.1 The Vector Form Factor of Pions

In particle physics, one of the best to study reactions of elementary particles, is the collision between an electron (e^-) and it's anti-particle, the positron (e^+). When these two particle collide, they annihilate each other and produce new types of particles. In these experiments very precise measurements can be taken and as such, be a very valuable base of empirical data of the Standard model of physics. A central point of study, of these electron-positron-collisions has been the anomalous magnetic moment g-2 of the muon. The anomaly of this number comes from the fact, that the measured data differs to the theoretical model by quite a large margin. As such it could be the source of exciting discoveries. The theoretical value of the g-2 momentum relies on data from the aforementioned collisions, which is used to reconstruct the so called hadronic vacuum polarization. The hadronic vacuum polarization itself comes from the hadronic final states. About 70% of the contribution to the g-2 momentum comes from the annihilation of an electron and a positron into two pions. The probability of this happening is dependent on the energy of the two particles. The strong interaction between these two pions is given by the so called pion vector form factor (VFF, F_π^V). To obtain the pion VFF, we start with a classical damped driven oscillator, which can be described as

$$\frac{d^2x}{dt^2} + \gamma\frac{dx}{dt} + \omega_0^2 x = A\cos(\omega t) \tag{1}$$

that has a solution $x(t) = K\cos(\omega t - \Phi)$ The coefficient K describes the amplitude, which is a function of the natural frequence ω_0 of the oscillator, the damping coefficient γ, the driving frequency ω and the amplitude A:

$$K^2 = \frac{A^2}{(\omega^2 - \omega_0^2)^2 + \gamma^2\omega^2} \tag{2}$$

If now $\omega_0^2 > \frac{\gamma^2}{2}$, the peak of the ampöitudes appears at the resonance frequency
$\omega = \omega_R := \sqrt{\omega_0^2 - \frac{\gamma^2}{2}}$. This phenomenon can be transferred to the pion VFF, as a very similar effect happens

there. An unstable particle, the ρ meson with a mass of $M_\rho = 0.77GeV$ and a decay width $\Gamma_\rho = 0.15GeV$, acts as a resonance. Now the parameters need to be transcribed to the relativistic particle-physics context. The driving frequency is replaced by the invariant squared energy s, the resonance mass takes the place of the natural frequency and the decay width acts as the damping coefficient. As such, we obtain the Breit-Wigner form of the VFF:

$$|F_\pi^V(s)|^2_{BW,\rho} = \frac{M_\rho^4}{(M_\rho^2 - s)^2 + \Gamma_\rho^2 M_\rho^2} \tag{3}$$

This can also be written in complex form, as:

$$F_\pi^V(s)_{BW,\rho} = \frac{M_\rho^2}{M_\rho^2 - s - iM_\rho\Gamma_\rho} \tag{4}$$

Although representing the VFF quite well, it is still very simplistic and has to be modified. By implementing another resonance contribution, the Vector Form Factor can be brought into the form:

$$F_\pi^V(s)_{BW,\rho+\omega} = \frac{M_\rho^2}{M_\rho^2 - s - iM_\rho\Gamma_\rho(s)} \times (1 + \varepsilon_\omega \frac{s}{M_\omega^2 - s - iM_\omega\Gamma_\omega}) \times (1 + as + bs^2 + cs^3) \tag{5}$$

or the modulus given as:

$$|F_\pi^V(s)|^2 = \frac{M_\rho^4}{(M_\rho^2 - s)^2 + M_\rho^2\Gamma_\rho(s)^2} \times (1 + \varepsilon_\omega \frac{2s(M_\omega^2 - s)}{(M_\rho^2 - s)^2 + M_\rho^2\Gamma_\rho(s)^2} \times (1 + as + bs^2 + cs^3) \tag{6}$$

2.2 The D'Agostini bias

The D'Agostini bias was first introduced by Giuilo D'Agostini in 1994. It describes a problem with data-fits, when considering data with overall systematic errors, that share a uncertainty on the normalization factor. In such a situation, if the error matrix V of the data points is known, one would normally minimize the χ^2, which can be obtained by

$$\chi^2 = \vec{\Lambda}^T \cdot V^{-1} \cdot \vec{\Lambda} \tag{7}$$

In this formula, Λ denotes the vector between the values of the theoretical model and the measured ones. But, after carrying out such a fit, one often obtains results, which contradict expectations. For example, if we got the results $8.0 \pm 2\%$ and $8.5 \pm 2\%$, from a measurement, which share a 10% normalization error, if we minimized the χ^2 as described-with the matrix V estimated by the data, we would obtain the value 7.87 ± 0.81. This result should immediately take attention, as the result with the highest probability, lies outside the range of the measured values. This error also occurs in a situation, where data is taken from two or more independently conducted experiments, which are afflicted by an additional systematic normalization error, even though the dimensions of the error are not quite as severe as in the situation described before.

2.3 Iterative solution to the D'Agostini bias

The proposed solution of the D'Agostini bias, that avoids problems with

multiple experiments or quadraticity with the parameters, is as follows. The covariance matrix is constructed, not by the fit result, but by a fixed guessed value $y_0 = f(x, \vec{p}_0)$. For one experiment, we then obtain: The covariance matrix calculated by:

$$[Cov_{ij}]_{syst} = \zeta_{ij}^2 |F_p i^V(s_i)|^2 |F_p i^V(s_j)|^2 \tag{8}$$

can be wrtitten as:

$$[Cov(y_i,y_j)]_{syst} = \begin{pmatrix} rr\zeta^2 y_0^2 & \zeta^2 y_0^2 \\ \zeta^2 y_0^2 & \zeta^2 y_0^2 \end{pmatrix} = \begin{pmatrix} rr\zeta^2 f(x_1,\vec{p}_0)^2 & \zeta^2 f(x_1,\vec{p}_0) f(x_2,\vec{p}_0) \\ \zeta^2 f(x_1,\vec{p}_0) f(x_2,\vec{p}_0) & \zeta^2 f(x_2,\vec{p}_0)^2 \end{pmatrix} \tag{9}$$

or for two independent experiments:

$$[Cov(y_i,y_j)]_{syst} = \begin{pmatrix} rr\zeta_1^2 y_0^2 & 0 \\ 0 & \zeta_2^2 y_0^2 \end{pmatrix} = \begin{pmatrix} rr\zeta_1^2 f(x_1,\vec{p}_0)^2 & 0 \\ 0 & \zeta_2^2 f(x_2,\vec{p}_0)^2 \end{pmatrix} \tag{10}$$

This conserves the quadraticity of the error functions parameters, that appear in the linear model. After solving this model, one obtains new estimates for the parameters $\vec{p}$. With these solutions a new systematic covariance can be constructed and with each iteration, the new values are used for the next construction. This iterative solutions is called the "$t_0 - method$".

2.4 Code Structure

Here the logical structure of the code is shown.

- Construct statistical covariance matrix
- Construct Jacobi matrix of model function in terms of the parameters
- Guess initial parameters $\vec{p}_0$
- Iterate
  - Construct System covariance matrix with $\vec{p}_i$
  - Fill Jacobi matrix with $\vec{p}_i$ which is the Design Matrix
  - Calculate step $\delta\vec{p}_i$
  - update initial parameters

$$\vec{p}_{i+1} = \vec{p}_i + \alpha \cdot \delta\vec{p}_i \quad \alpha = 0.1$$

- Calculate errors
- Calculate χ^2_{min}

The guess used for all fits was determined by standard least-square fit provided by `scipy`.

$$\vec{p}_0 = \big(900,\ 200,\ 810,\ 40,\ 20,\ -1000,\ 840,\ 1550\big) \tag{11}$$

The code can be viewed and/or downloaded from here [1] (including the calculation of the guess parameters).

3 Findings

In this section the results with consideration of the D'Agostini are shown. Furthermore the findings with fits of two experiments together (6 combinations of two) and also a fit with all experiments are shown. In the end of the section the fitted parameters are compared with the literature values. The plots of the given data and their fits can be found in Section 3.5.

3.1 Single Experiment Fits under consideration of the D'Agostini bias

In this section the data is fitted under consideration of the D'Agostini bias of all experiments separately is shown.

Table 1: Results of all experiment data fitted separately

$\vec{p}$	SND	CMD2	KLOE	BABAR
M_ρ [MeV]	772.72 ± 0.59	773.93 ± 0.67	773.91 ± 0.25	773.33 ± 0.43
Γ_ρ [MeV]	149.53 ± 1.15	147.67 ± 1.32	149.72 ± 0.37	149.19 ± 0.81
M_ω [MeV]	781.94 ± 0.09	782.32 ± 0.07	782.44 ± 0.11	782.18 ± 0.07
Γ_ω [MeV]	8.55 ± 0.33	8.65 ± 0.44	9.66 ± 0.33	8.17 ± 0.16
ε_ω []	2.02 ± 0.09	1.92 ± 0.12	2.07 ± 0.05	1.95 ± 0.03
χ^2_{min}/dof	1.001	1.054	1.443	1.031
p-value	0.530	0.395	0.001	0.377

3.2 Multi Experiment Fits under consideration of the D'Agostini bias

In this section the data is fitted considering the D'Agostini bias, first the data of two experiments together then the data of all experiments is shown.

Table 2: Results of data fits of experimental data fitted in pairs

$\vec{p}$	SND-CMD2	SND-KLOE	SND-BABAR
M_ρ[MeV]	772.72 ± 0.42	773.92 ± 0.23	773.17 ± 0.36
Γ_ρ[MeV]	149.53 ± 0.81	149.42 ± 0.35	149.70 ± 0.64
M_ω [MeV]	781.95 ± 0.07	782.39 ± 0.07	782.07 ± 0.06
Γ_ω [MeV]	8.56 ± 0.24	9.42 ± 0.20	8.27 ± 0.13
ε_ω[]	2.02 ± 0.07	2.07 ± 0.05	1.96 ± 0.03
χ^2_{min}/dof	0.904	1.839	0.945
p-value	0.754	0.001	0.763

Table 3: Results of data fits of experimental data fitted in pairs

$\vec{p}$	CMD2-KLOE	CMD2-BABAR	KLOE-BABAR
M_ρ [MeV]	773.92 ± 0.23	773.17 ± 0.36	773.66 ± 0.20
Γ_ρ [MeV]	149.42 ± 0.35	149.70 ± 0.64	$149.41\pm$
M_ω [MeV]	782.39 ± 0.07	782.07 ± 0.06	782.49 ± 0.06
Γ_ω [MeV]	9.42 ± 0.02	8.27 ± 0.13	8.98 ± 0.12
ε_ω []	2.07 ± 0.05	1.96 ± 0.03	1.98 ± 0.02
χ^2_{min}/dof	1.838	0.943	1.470
p-value	0.001	0.772	0.001

Table 4: Results of data fit of all experimental data fitted together

$\vec{p}$	Multi-Fit
M_ρ [MeV]	773.62 ± 0.18
Γ_ρ [MeV]	149.42 ± 0.29
M_ω [MeV]	782.36 ± 0.08
Γ_ω [MeV]	8.75 ± 0.08
ε_ω[]	1.96 ± 0.02
χ^2_{min}/dof	1.735
p-value	0.000

3.3 Litrature comparison

In this section the fitted parameters of CMD2-BABAR and the literature values[2] are compared. The reason why CMD2-BABAR was chosen, is that it has a value of χ^2_{min} close to 1.

Table 5: Result comparison with literature

$\vec{p}$	Literature	CMD2-BABAR	Relative error
M_ρ [MeV]	775.26 ± 0.25	773.17 ± 0.36	0.28%
Γ_ρ [MeV]	147.80 ± 0.90	149.70 ± 0.64	1.29%
M_ω [MeV]	782.65 ± 0.12	782.07 ± 0.06	0.08%
Γ_ω [MeV]	8.49 ± 0.08	8.27 ± 0.13	2.60%

3.4 Fitting with D'Agostini bias

Furthermore here we compare the results of the t0-method, with the incorrect method of multiplying the relative systematic uncertainties with the data provided, instead of calculating the systematic uncertainties in regards of the newly calculated parameters in every iteration. For demonstrational purposes only the single fitted experiments are pulled. Again we would like to reference the plots in Section 3.5. In the following table the results of the incorrect method applied to fitting are shown.

Table 6: Results of all experiment data wrongly fitted separately

$\vec{p}$	SND	CMD2	KLOE	BABAR
M_ρ [MeV]	772.21 ± 4.97	774.51 ± 2.59	774.09 ± 0.25	773.42 ± 0.54
Γ_ρ [MeV]	151.26 ± 13.11	145.18 ± 5.83	149.89 ± 0.84	149.29 ± 1.01
M_ω [MeV]	781.03 ± 0.47	782.09 ± 0.32	782.79 ± 0.18	782.18 ± 0.07
Γ_ω [MeV]	8.96 ± 2.25	8.58 ± 0.15	11.11 ± 0.70	8.18 ± 0.18
ε_ω[]	2.14 ± 0.68	1.85 ± 0.38	2.15 ± 0.08	1.94 ± 0.04
χ^2_{min}/dof	0.016	0.125	0.903	1.118
p-value	1.000	1.000	0.839	0.099

3.5 Plots

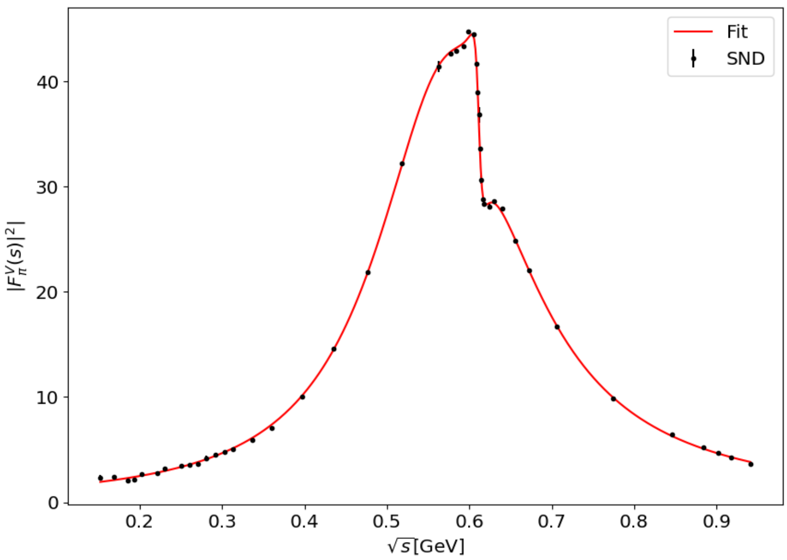

Figure 1: SND data fit

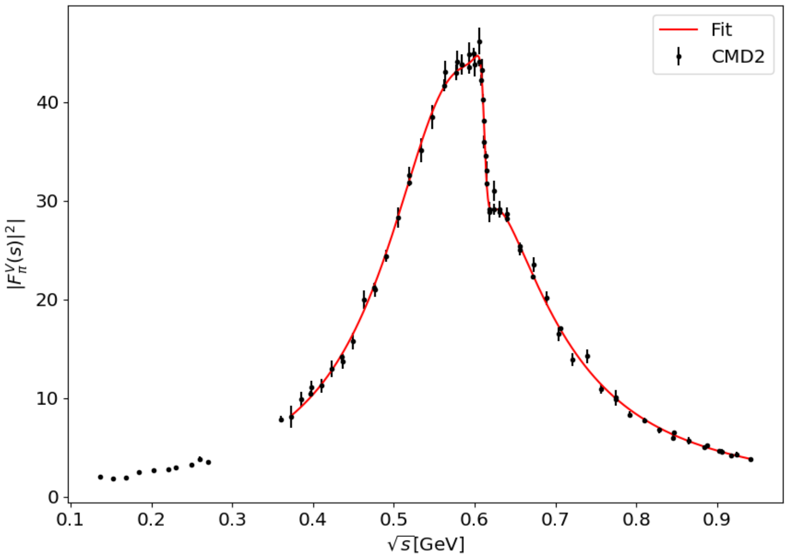

Figure 2: CMD2 data fit

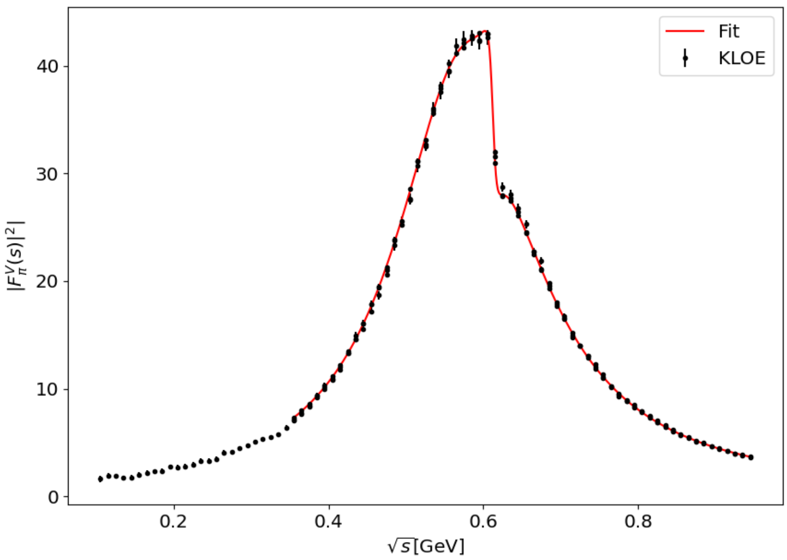

Figure 3: KLOE data fit

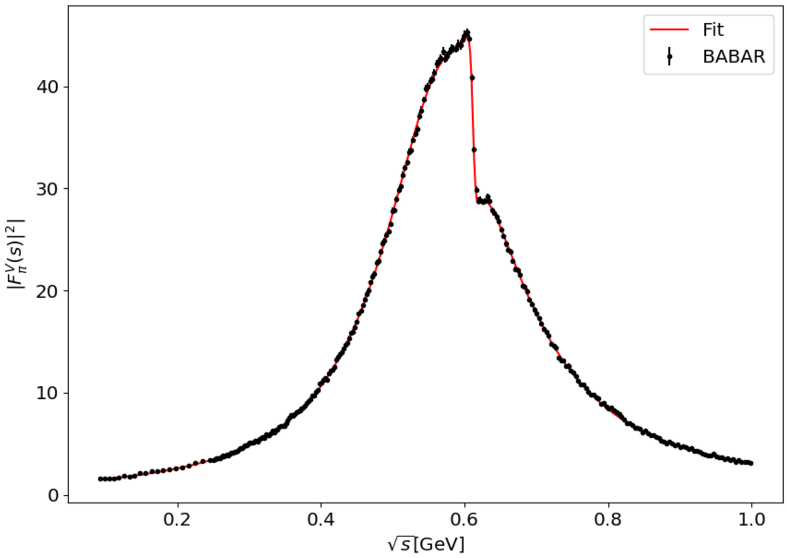

Figure 4: BABAR data fit

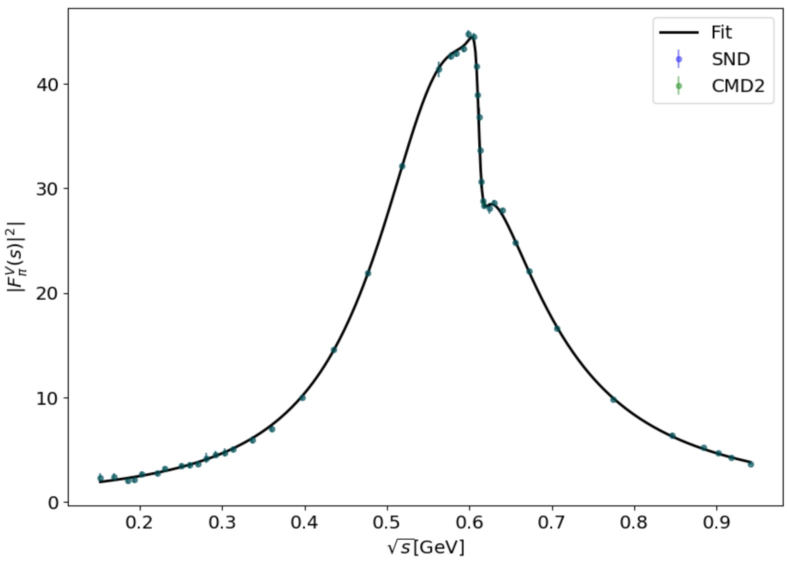

Figure 5: SND and CMD2 fitted togther

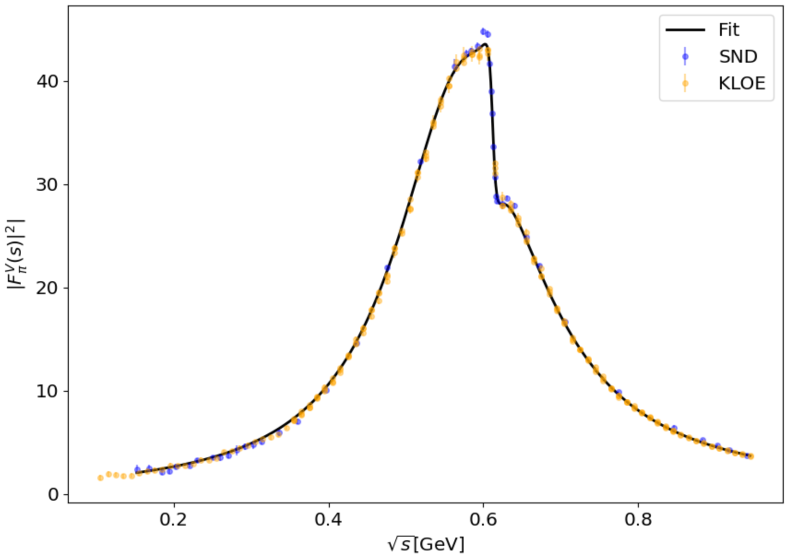

Figure 6: SND and KLOE fitted togther

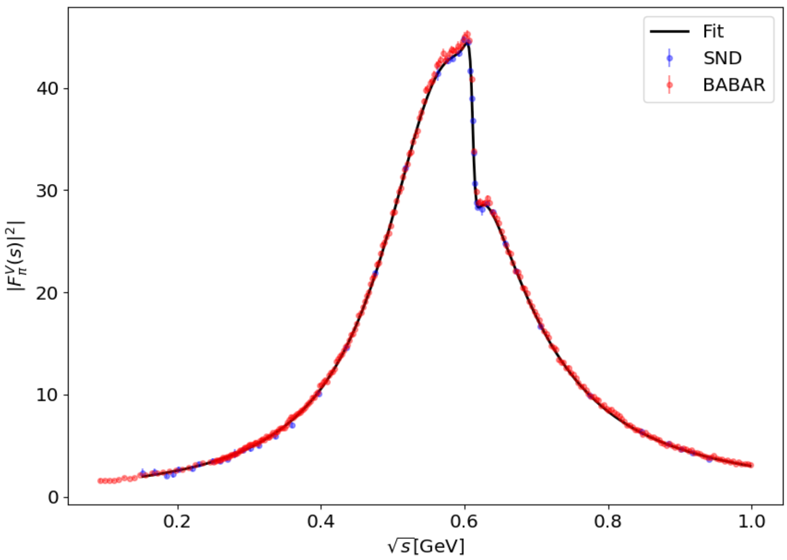

Figure 7: SND and BABAR fitted togther

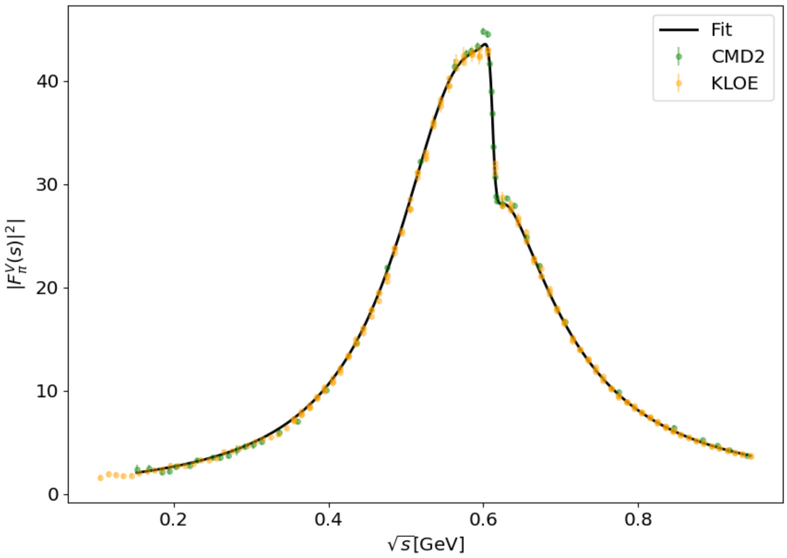

Figure 8: CMD2 and KLOE fitted togther

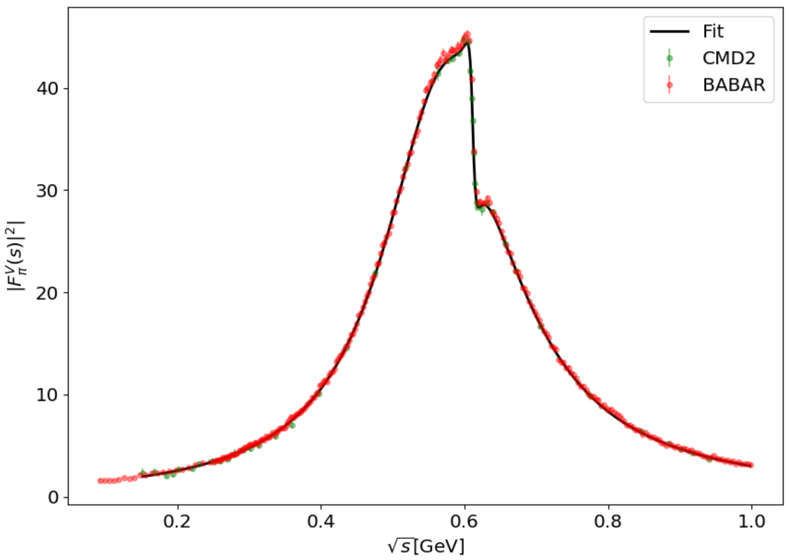

Figure 9: CMD2 and BABAR fitted togther

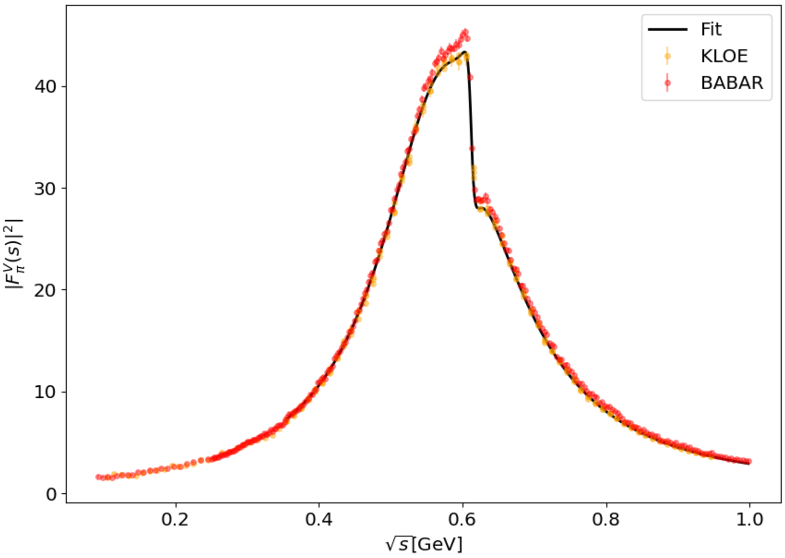

Figure 10: KLOE and BABAR fitted togther

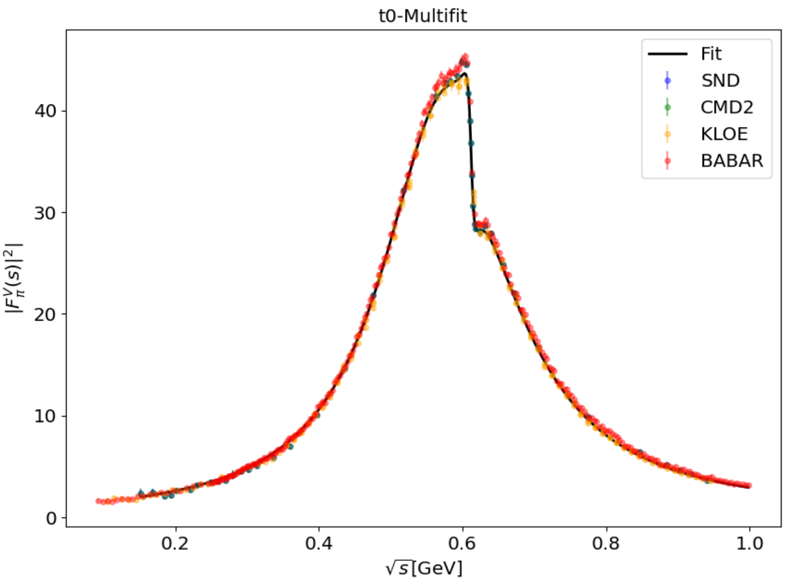

Figure 11: SND, CMD, KLOE and BABAR fitted togther

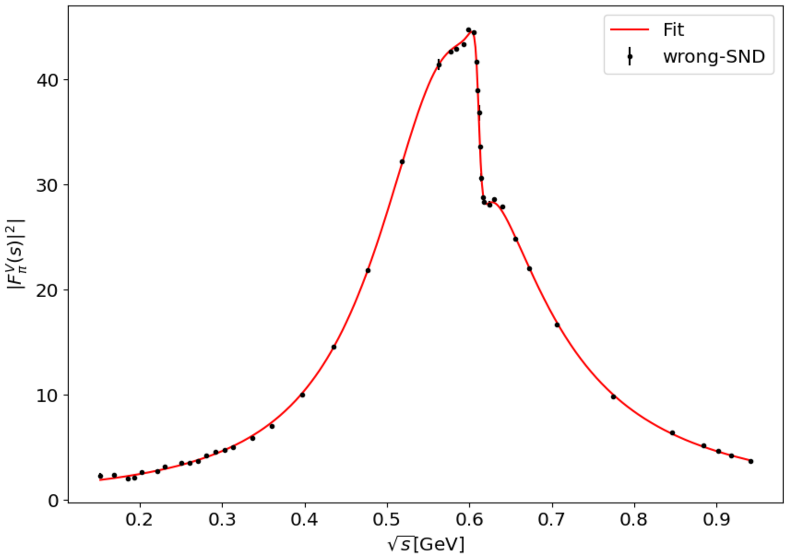

Figure 12: Wrong method, SND data fit

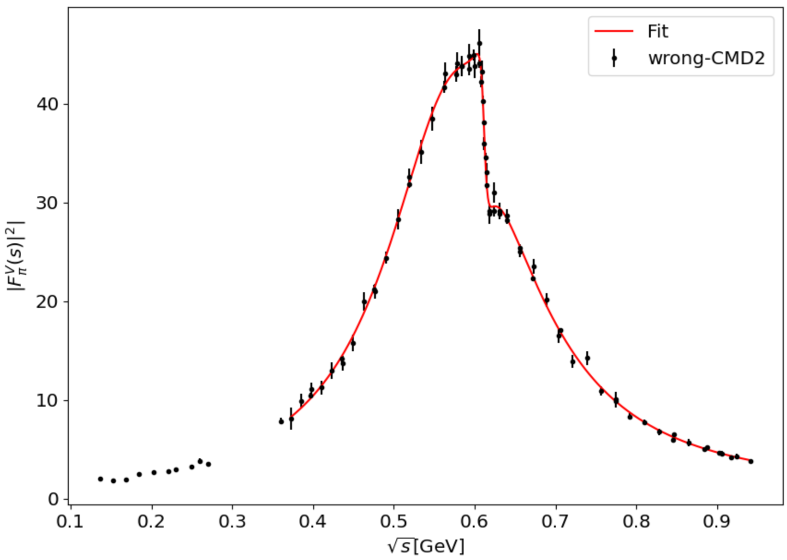

Figure 13: Wrong method, CMD2 data fit

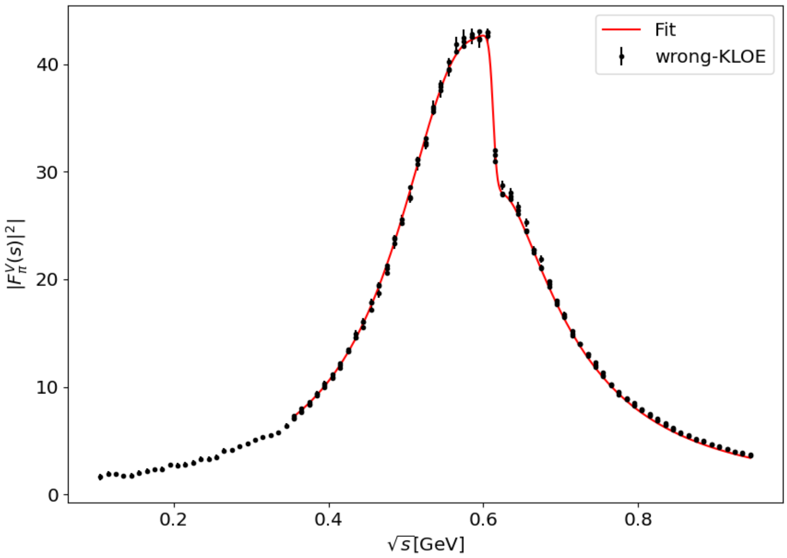

Figure 14: Wrong method, KLOE data fit

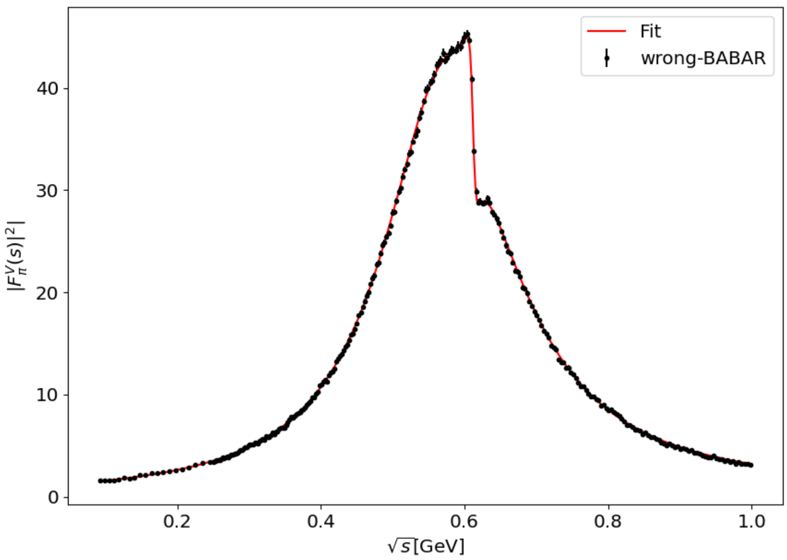

Figure 15: Wrong method, BABAR data fit

4 Conclusion

Under consideration of the D'Agostini bias in the calculations we arrive very close the the literature values in table 5. Furthermore the parameters in table 1 all make the χ^2_{min}/dof value converge to 1, making the goodnes of the fit very good. When fitting two experiments together the results according to the χ^2_{min}/dof value are only good for SND-CMD2, SND-BABAR and CMD2-BABAR, table 2 and 3. Taking this into account we need to choose the experiments fitted together very carefully to arrive at good results, fitting all experiments together for instance does't provide a good χ^2_{min}/dof value, table 4. Furthermore looking at the parameter fits without consideration of the D'Agostini bias we can conclude that the fits are obviously missing something, table 6.

References

[1] *Git Instance, Implementation of the T0-Method*. 2021. URL: `git://popovic.xyz/tprak.git` (visited on 04/17/2021).

[2] Particle Data Group. *Particle Data Group*. 2020. URL: `https://pdg.lbl.gov/` (visited on 04/17/2021).