

Manuscript Number: CBM-D-16-00437R2

Title: A CONWAY-MAXWELL-POISSON (CMP) MODEL TO ADDRESS DATA DISPERSION ON
POSITRON EMISSION TOMOGRAPHY

Article Type: Full Length Article

Keywords: positron emission tomography (PET), Conway-Maxwell-Poisson
(CMP) distribution, maximum likelihood (ML) estimation, sinograms,
Poisson statistic deviation

Corresponding Author: Dr. Maria Filomena Santarelli, Ph.D.

Corresponding Author's Institution: National Research Council (CNR)

First Author: Maria Filomena Santarelli, Ph.D.

Order of Authors: Maria Filomena Santarelli, Ph.D.; Daniele Della Latta,
BSc; Michele Scipioni, MSc; Vincenzo Positano, MSc; Luigi Landini, Prof

Abstract: Positron emission tomography (PET) in medicine exploits the properties of positron-emitting unstable nuclei. The pairs of γ - rays emitted after annihilation are revealed by coincidence detectors and stored as projections in a sinogram. It is well known that radioactive decay follows a Poisson distribution; however, deviation from Poisson statistics occurs on PET projection data prior to reconstruction due to physical effects, measurement errors, correction of deadtime, scatter, and random coincidences. A model that describes the statistical behaviour of measured and corrected PET data can aid in understanding the statistical nature of the data: it is a prerequisite to develop efficient reconstruction and processing methods and to reduce noise. The deviation from Poisson statistics in PET data could be described by the Conway-Maxwell-Poisson (CMP) distribution model, which is characterized by the centering parameter λ and the dispersion parameter ν , the latter quantifying the deviation from a Poisson distribution model. In particular, the parameter ν allows quantifying over-dispersion ($\nu < 1$) or under-dispersion ($\nu > 1$) of data. A simple and efficient method for λ and ν parameters estimation is introduced and assessed using Monte Carlo simulation for a wide range of activity values. The application of the method to simulated and experimental PET phantom data demonstrated that the CMP distribution parameters could detect deviation from the Poisson distribution both in raw and corrected PET data. It may be usefully implemented in image reconstruction algorithms and quantitative PET data analysis, especially in low counting emission data, as in dynamic PET data, where the method demonstrated the best accuracy.

Conflict of Interest Statement

None declared

Pisa, 04/08/2016

Manuscript # CBM-D-16-00437: "A Conway-Maxwell-Poisson (CMP) Model to Address Data Dispersion On Positron Emission Tomography"

Authors: M.F. Santarelli, D. Della Latta, M. Scipioni, V. Positano, L. Landini

Dear Editor,

we are submitting the manuscript, after English grammatical structure and phraseology corrections, as required. An US-English mother-tongue linguist made them.

All the changes/adjunctions performed on the manuscript are highlighted in the text.

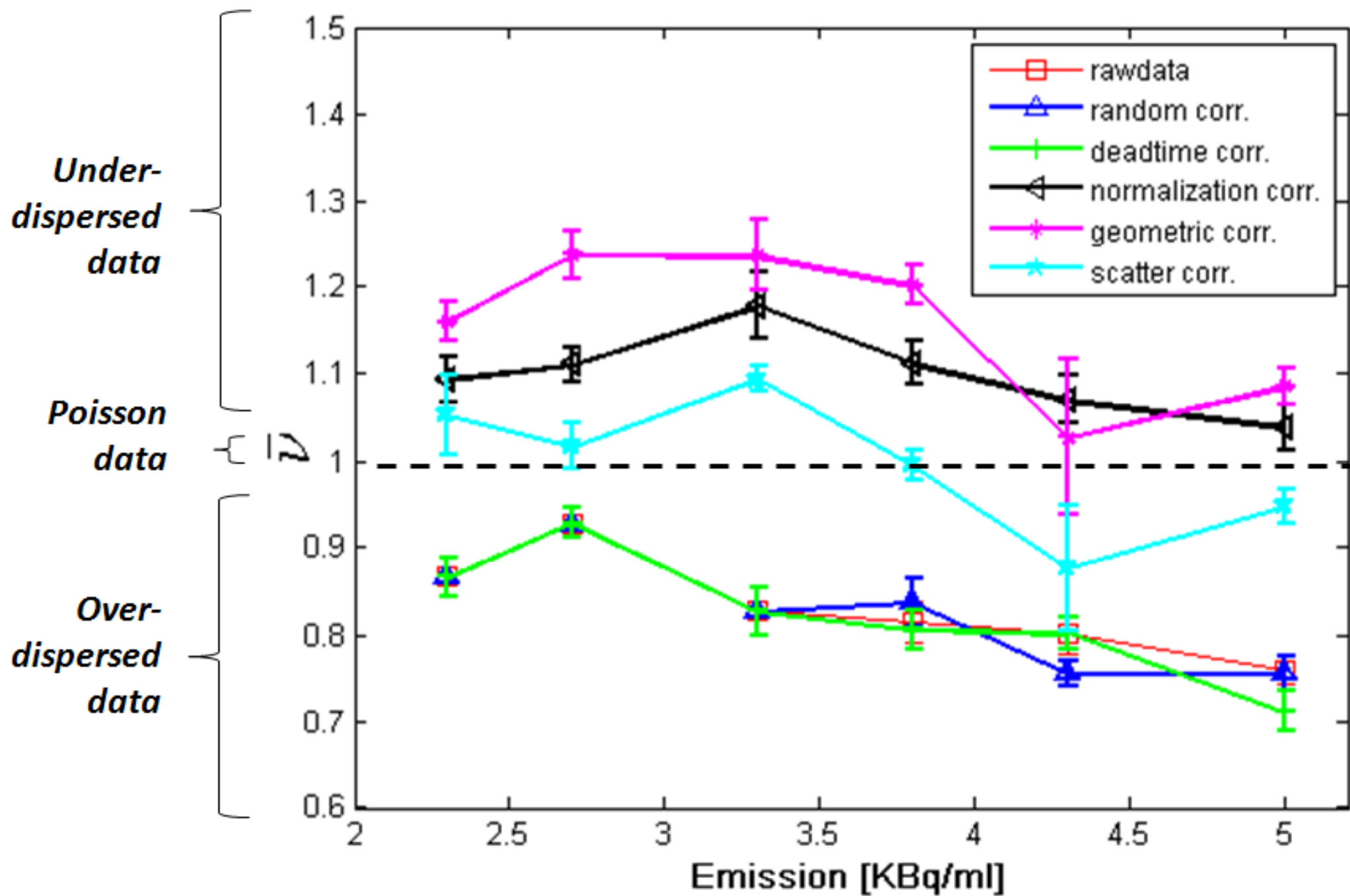
Best regards

Maria Filomena Santarelli and Co-authors

*Revision Notes

We are submitting the manuscript, after English grammatical structure and phraseology corrections, as required by the Editor.
An US-English mother-tongue linguist made corrections.

PET data: experimental phantom



Highlights

- Deviation from Poisson statistics occurs on PET projection data due to physical effects, measurement errors, correction of deadtime, scatter and random coincidences.
- The Conway-Maxwell-Poisson (CMP) distribution model allows to quantify over-, equi and under-dispersion of data.
- The performance of CMP distribution for PET dataset has been proved using both simulation and measured data.

A CONWAY-MAXWELL-POISSON (CMP) MODEL TO ADDRESS DATA DISPERSION ON POSITRON EMISSION TOMOGRAPHY

Maria Filomena Santarelli^{*a,b}, Daniele Della Latta^b, Michele Scipioni^c, Vincenzo Positano^b, Luigi Landini^{c,b,a}

a. Institute of Clinical Physiology, National Research Council, via Moruzzi 1, 56124 Pisa, Italy;

b. Fondazione CNR-Regione Toscana "G. Monasterio", via Moruzzi, 1, 56124 Pisa, Italy;

c. Department. of Information Engineering: EIT, University of Pisa, via Diotisalvi 2, 56126 Pisa, Italy.

* Correspondence to: Institute of Clinical Physiology, National Research Council, via Moruzzi 1, 56124 Pisa, Italy Tel.: +39 050 3152614; fax: +39 050 3152166; e-mail: santarel@ifc.cnr.it (M.F. Santarelli)

ABSTRACT

Positron emission tomography (PET) in medicine exploits the properties of positron-emitting **unstable** nuclei. The pairs of γ - rays emitted after annihilation are revealed by coincidence detectors and stored as projections in a sinogram. It is well known that radioactive decay follows a Poisson distribution; however, deviation from Poisson statistics occurs on PET projection data prior to reconstruction due to physical effects, measurement errors, correction of deadtime, scatter, and random coincidences. A model that describes the statistical behaviour of measured and corrected PET data can aid in understanding the statistical nature of the data: it is a prerequisite to **develop** efficient reconstruction and processing methods and to reduce noise.

The deviation from Poisson statistics in PET data could be described by the Conway-Maxwell-Poisson (CMP) distribution model, which is characterized by the centring parameter λ and the dispersion parameter ν , the latter quantifying the deviation from a Poisson distribution model. In particular, the parameter ν allows **quantifying** over-dispersion ($\nu < 1$) or under-dispersion ($\nu > 1$) of data. A simple and efficient method for λ and ν parameters estimation is introduced and assessed using Monte Carlo simulation for a wide range of activity values.

The application of the method to simulated and experimental PET phantom data demonstrated that the CMP distribution parameters **could** detect deviation from the Poisson distribution both in raw and corrected PET data. It may be usefully implemented in image reconstruction algorithms **and** quantitative PET data analysis, especially in low counting emission data, as in dynamic PET data, where the method demonstrated the best accuracy.

KEYWORD: positron emission tomography (PET), Conway-Maxwell-Poisson (CMP) distribution, maximum likelihood (ML) estimation, sinograms, Poisson statistic deviation.

Nomenclature	
CMP	Comway-Maxwell-Poisson
ML	maximum likelihood
ML-EM	maximum likelihood expectation maximization
NB	negative binomial
PET	positron emission tomography
WLS	weighted least square
Se	standard error
$\%bias$	bias index in percent
$E[X]$	mean of the CMP process
$Var[X]$	variance of the CMP process
D	dispersion index
n	sample size
N	number of repetitions of the same experiment
Nv	number of vectors, i.e. central line of the sinograms, used in the analysis of experimental phantom data
Nl	number of sequenced vectors, used in the analysis of experimental phantom data
λ	centring parameter in the CMP distribution; it is related to the observations mean
ν	shape parameter in the CMP distribution; it is related to data over-dispersion and under-dispersion
$\hat{\lambda}$	estimated centring parameter over the sample size n
$\bar{\lambda}$	mean of the $\hat{\lambda}$ value, over N repetitions
$\hat{\nu}$	estimated dispersion (shape) parameter over the sample size n
$\bar{\nu}$	mean of the $\hat{\nu}$ value, over N repetitions
θ	dummy variable, stands for ν and λ
$\hat{\theta}_i$	estimated value of θ at each repetition

1. Introduction

Today numerical and computational methods are becoming increasingly frequent in biology and medicine to gain new insights into the emerging research trends [1-4]. Among those benefiting from the numerical and computational methods, medical imaging plays an integral role, especially in the image formation process. In particular, in PET imaging such methods are crucial to grasp both the biological and the technological aspects of image acquisition and reconstruction processes. Actually, in positron emission tomography (PET) a tracer containing a radioactive isotope is introduced into the body, and it emits positrons while being involved in metabolic processes. Due to annihilation between a positron and an electron, a pair of γ -rays is emitted in opposite directions along a line of response (LOR) at each spatial coordinate; the γ -ray pair is revealed by a coincidence detectors pair placed in an array of detector units surrounding the body. Emission of γ -rays follows a Poisson law [5]. In this paper, numerical and computational methods are used to investigate deviation from Poisson statistics occurring in PET projection data before (or during) reconstruction.

A known source of deviation from Poisson model is related to errors introduced in correcting some unwanted physical and acquisition effects, such as scattering errors, accidental errors, positron range, deadtime, detectors non-collinearity, variation in detection-pair sensitivity, and others. [6-9]. Scattered coincidences are due to one or multiple Compton scatterings of the detected photons; as a result, wrong LORs are obtained, which cause a reduced contrast and an overestimation of the tracer activity [10]. Random (or false) coincidences occur when photons detected within the same time window are not generated from the same annihilation event. Random events cause a reduction of image contrast and a bias in the measured activity due to an increased level of background activity [11]. Both scattering and random coincidences are not wanted events, which need a correction for obtaining exact emission data; however, such corrections cause deviation from Poisson statistics [12]. A further source of deviation from Poisson law is the deadtime, i.e. the instrumental insensitivity to detectable events due to electronic or mechanical reset time. Especially in high counting rate, some events can be missed, and the detected rate is lower than the true one [13]. Reduction in the mean value is a potential cause of the deviation from the Poisson model. The detectors non-collinearity and the variation in detection-pair sensitivity are two scanner-related factors that cause an incorrect emission evaluation, and need correction [11]. The sensitivity of the LOR about the mean is affected both by the geometry of the detectors and the LOR position; hence,

a geometric correction is needed to get a correct detection. Besides such geometric effects, the block detectors themselves vary in efficiency, as the photomultiplier tubes gains are not uniform (and may change in time), and the crystals are not all identical; then a normalization operation is required. In commercial PET scanners, in addition to the stochastic nature of the photon-counting, the intrinsic thermal and electronic noise should be accounted for. Such sources of artefacts are usually described by a Poisson model corrupted with additive Gaussian noise [14]. When deviation from Poisson model is notable, Poisson assumption cannot be applied anymore. Such trade-off could be suitably steered by the knowledge of the true statistical model of sinogram data. Unfortunately, the degree of deviation from Poisson model remains unknown for the lack of a reliable method that can quantify such deviations.

Using a fully 3D maximum likelihood expectation maximization (ML-EM) reconstruction algorithm, two correction methods are currently used. In one method measured data are pre-corrected before applying the reconstruction algorithm. We know that pre-correction for random and scatter coincidences may result in deviation from Poisson assumption, depending on the entity of correction. It compromises the estimation of the emission density based on ML reconstruction [15]. The second correction method incorporates correction models into the iterative algorithm, thus preserving the Poisson nature of the data [16]. Such solution is not immune to errors due to lack of knowledge about the true statistics of sinogram data. Moreover, such approach produces bias and higher variability in applications where images are reconstructed from a relatively small number of counts [17-20]. This issue is an important aspect to take into account, especially for cold regions and/or dynamic PET data reconstruction; actually, bias and variability of reconstructed data can lead to an incorrect activity evaluation and inexact kinetic parameters estimate. Without any doubts, an accurate evaluation of the deviation of the measured PET data from Poisson statistics could lead to a more appropriate correction, reducing bias and variability.

In the literature, a few classes of discrete probability models were proposed and studied as alternatives to Poisson model in measured data. Sometimes, the measured data describe an over-dispersion or an under-dispersion as a deviation from the Poisson model [21]. In PET data analysis, the Negative Binomial (NB) distribution model has been proposed in monitoring the over-dispersion on both uncorrected or corrected (random and scatter coincidence correction) sinograms. [22,23]. In such papers, the authors highlighted the importance of an accurate modelling of PET data to develop efficient reconstruction algorithms. To account for over and under-dispersion

simultaneously, the Conway-Maxwell-Poisson (CMP) distribution model can be suggested. It is worth to note that the CMP distribution model has been studied extensively [24-28]. Applications of the CMP distribution model range from Biology [29] to Transportation [30], Marketing, and eCommerce [31,32]. However, to our knowledge, the use of CMP distribution model to describe PET data statistics remains unexplored. So, the application of the CMP model to the analysis of PET sinogram data is novel. The relevant benefits are twofold. Firstly, the CMP model allows to fully characterize the statistical behavior of PET sinogram data, by accounting for both uncorrected and corrected data, thus making possible the inclusion in the correction model of random, deadtime, normalization, geometric, and scatter coincidence effects. Secondly, the proposed approach may prelude to the application of the CMP model in the wide field of high energy medical imaging, in the situation where deviations from the conventionally used Poisson distribution model may happen.

This paper describes a method that uses the CMP model to assess the statistics of PET projections, collected into sinograms, in a wide range of emission rates. We focused on uncorrected and corrected (for random, deadtime, normalization, geometric, and scatter coincidence effects) sinograms, due to their importance in the quantitative analysis of PET data. A method that accurately defines the statistics of PET measured data, also in the conditions of deviations from Poisson statistics, could have important implications for the quantitative evaluation of PET data.

2. Materials and Methods

2.1 The Conway-Maxwell-Poisson (CMP) distribution model

2.1.1 The CMP probability function and its moments

The importance of CMP distribution model is due to its ability to model count data with different degrees of over- and under-dispersion. This distribution is a two-parameters extension of the Poisson distribution. Its probability density function is [24]:

$$P(X = x; \lambda, \nu) = \frac{\lambda^x}{(x!)^\nu} [Z(\lambda, \nu)]^{-1}, \quad \lambda > 0, \nu \geq 0, x = 0, 1, \dots$$

(1)

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}$$

where x denotes the amplitude data, λ is a centring parameter related to the observations mean, ν is the shape parameter of the CMP distribution, and Z is a normalizing constant.

The ν parameter outlines deviations from Poisson distribution: $\nu > 1$ corresponds to under-dispersed data, $\nu < 1$ to over-dispersed data, and $\nu = 1$ to equi-dispersed (Poisson) data. As $\nu = 1$ the mean value becomes equal to λ (the Poisson parameter).

Previous relationship (1) allows for the following representation as a non-linear decrease in ratio of successive probabilities [33]:

$$\frac{P(X = x-1)}{P(X = x)} = \frac{x^\nu}{\lambda} \quad (2)$$

Such formulation enables us to use **the** weighted least squares (WLS) method to estimate the CMP parameters. The λ and ν parameters can be used to derive an approximation of the mean, $E[X]$, and the variance, $\text{var}[X]$, of the CMP process according to the following relationships [33]:

$$E[X] = \lambda^{1/\nu} + \frac{1}{2\nu} - \frac{1}{2} \quad (3)$$

$$\text{var}[X] = \frac{1}{\nu} \lambda^{1/\nu} \quad (4)$$

It follows that, for $\nu=1$, $E[X] = \lambda$ and $\text{var}[X] = \lambda$, which corresponds to a Poisson process.

2.1.2 CMP parameters estimation method

The λ and ν parameters can be estimated using the approach described in [33]. The approach is based on eq. (2), relating the ratio of successive CMP probabilities to a non-linear relationship with

the λ and ν parameters. The following regression equation is found, taking the logarithm of both sides of eq.(2) [26]:

$$\log(p(x-1)/ p(x)) = -\log(\lambda) + \nu \log(x) \quad (5)$$

where $p(x)$ denotes $P(X=x)$. In actual situations, the ratio of probabilities on the left side of eq.(5) can be substituted with the relative frequency of occurrence of $(x-1)$ and x . The amplitude x is divided into a number of intervals, or classes, and for each class, the frequency of occurrence is evaluated. The number of classes is chosen in order to avoid zero counts in every class.

The estimation process can be split into two steps: 1) the term on the left side is plotted against the $\log(x)$ for every value of x ; 2) the best linear fitting is evaluated according to the WLS method to extract the CMP parameters. In the ideal CMP model, the data lie in a straight line, with angular coefficient ν and intercept $\log(\lambda)$. In the Poisson model ($\lambda=1, \nu=1$) the intercept would be equal to zero and the angular coefficient would be equal to one.

The percent-bias (*%bias*) and standard error (*se*) were used as performance indices of λ and ν parameters estimator. For each parameter, the *%bias* is defined as the difference between the expected and the true value. When *%bias* is zero, the estimator is unbiased. The *%bias* index is defined as follows[34]:

$$\%bias = \frac{1}{N} \sum_{i=1}^N \hat{\theta}_i - \theta \quad (6)$$

where θ is a dummy variable and it stands for ν and λ ; the mean computation is performed over N repetitions of the same experiment; $\hat{\theta}_i$ is estimated at each repetition over a sample of size n and θ is the true value.

The 'se' index is defined as the ratio between the standard deviation (*std*) of $\hat{\theta}_i$ ($i=1,..N$) estimates and the square root of the sample size[35]: $se = std/\sqrt{N}$.

The dispersion index, or Fano factor, D is defined as follows [35,36]:

$$D = \text{var}[X | \hat{\lambda}, \hat{\nu}] / E[X | \hat{\lambda}, \hat{\nu}] \quad (7)$$

with $E[X | \hat{\lambda}, \hat{\nu}]$ and $\text{var}[X | \hat{\lambda}, \hat{\nu}]$ calculated according to eq. (3) and eq. (4). It is an index which gives information about the over- or under-dispersion of data. The Poisson process is characterized by $D = 1$.

2.1.3 Data homoscedasticity

The ordinary regression model herein used is based on the assumption of homoscedasticity of data amplitude distributed on **the frequency** of occurrence classes. Because, as demonstrated in [33], such assumption is violated in the CMP distributed data (the variance of the dependent variable is not constant), we expect implications on parameters estimation accuracy. Firstly, the CMP plot is constructed plotting the logarithm of the ratio of successive frequencies ($\log(p(x-1)/p(x))$ vs $\log(x)$) for various combinations of n , λ and ν . On such a plot we evaluated the regression parameters over the full amplitude range of $\log(x)$. After that, we progressively reduced such interval symmetrically with respect to the amplitude of the maximum frequency of occurrence (i.e. the highest $P(X=x)$). For each amplitude **range**, the *std* of the $\hat{\nu}$ and $\hat{\lambda}$ estimates was evaluated. The amplitude interval of $\log(x)$ with an almost constant *std*, at each estimated parameters, was assumed as the best compromise for homoscedasticity, and it was used in the following simulations. With the constraint defined above, the parameters were estimated by fitting a linear regression to eq.(5).

2.1.4 Characterization of the CMP parameters by simulation

To test the estimation accuracy of the CMP-WLS model, we simulated a dataset characterized by a CMP distribution with known regression parameters. Such parameters correspond to a range of mean and variance values typical of clinical PET data. We then estimated the *%bias* and the *se* of the estimates using Monte Carlo simulation.

The Monte Carlo simulation was implemented according to the following steps:

1. Generation of repeated CMP data, combining the following parameters: sample size, n , ranging from 500 to 5000; true centring parameter, λ , ranging from 1 to 20; ν value, ranging from 0.75 to 1.5. The excursion of the ν parameter allows **simulating** data with over-, equi- and under-dispersion encountered in PET sinogram data. In particular, the variance ranges from $0.1 * \lambda$ to $10 * \lambda$.

The **R** software [37] with compGLM package was used for data generation.

2. For each combination of n , λ and V , the following estimations were performed:

- \hat{V} estimation over the sample size n for three λ values using eq. (5); each estimation was repeated $N = 50$ and 100 times;
- $\hat{\lambda}$ estimation over the sample size n for three V values using eq. (5); each estimation was repeated $N = 50$ and 100 times;
- evaluation of the mean \bar{v} over the \hat{v} values previously estimated;
- evaluation of the mean $\bar{\lambda}$ over the $\hat{\lambda}$ values previously estimated;
- evaluation of *%bias* and $\pm se$ of \hat{v} and $\hat{\lambda}$;

Parameters estimation and analysis were accomplished by using Matlab (The MathWorks, Inc., Natick, Massachusetts, United States) software.

2.2 Phantom data acquisition and analysis

A cylindrical phantom, with a radius of 12 cm and height of 15 cm, was filled with 18F-FDG and water solution (34MBq of total activity at the beginning of the experiment). The phantom was put inside a PET/CT (General Electric Discovery) tomograph, exactly in the middle of the bore, with its long axis parallel to the bore's axis.

The experiment involved the acquisition of 47 slices of 2D PET data projections covering the phantom, for six different times, with a time interval of about 15-20 minutes. Each volumetric data was acquired in a total time of 300 sec. Due to the emission decay, the total activity at the beginning of each volumetric acquisition was reduced consequently; so, the total emission at the beginning of each time acquisition ranged from 34MBq to 15MBq.

For each time interval, the 2D sinograms consisted of 315x367 values for 47 slices; the following volumetric sinograms were stored:

- raw-data, without any correction, and:
 - o after correction for random counts
 - o after correction for deadtime

- after normalization
- after geometric correction
- after correction for scattering

The sequence of corrections was applied according to a protocol implemented in a commercial scanner.

On each sinogram, its central line was selected and organized in a vector of 315 points. For a homogeneous circular object, the central line of the sinogram is relevant to the sum of the phantom contributions situated along its diameter, at each angle. A total of 47 vectors were available for the analysis. From these, 11 vectors were removed (they are relevant to 6 slices on one side and 5 slices on the other side of the phantom), leaving $nv=36$ vectors for the analysis. The Region-Of-Interest (ROI) for the regression analysis was realized by sequencing more vectors to increase the estimation accuracy: the ROI is constituted of $n = 315*nl$ data points, where nl is the number of sequenced vectors. The optimal nl value was determined according to simulation results.

The estimation of the \hat{V} and $\hat{\lambda}$ parameters was then performed on the selected ROI. In particular, for each volumetric data, the estimated parameters were averaged over $N = nv/nl=36/nl$ independent ROIs, each one consisting of n data points covering the middle of the phantom; the averaging can be done thanks to the homogeneity of the phantom. Such procedure was repeated for each corrected and not-corrected sinogram, and for each emission dose.

3. Results

3.1 CMP parameters estimation on simulated data

This section deals with the assessment of the performance of the proposed CMP parameters estimation method, using Monte Carlo simulation. Firstly, the amplitude interval of $\log(x)$ with a constant spread on the estimated parameters was identified. In Figure 1 four typical trends (plus the error bars (se)) of the logarithm of the ratios of successive frequencies ($\log(p(x-1)/p(x))$ of eq. (5)) vs $\log(x)$, for four λ values ($\lambda = 5, 10, 15, 20$), $n=1000$, and $V = 1$, are shown. Such plots highlight the spread (i.e. the se) of the dependent variable as a function of $\log(x)$: at the extremes, the spread

is larger than in the middle of $\log(x)$. So, for each λ value, an interval of $\log(x)$ data with almost constant se was identified to be used in the following estimates. Then, the accuracy on V and λ estimation was assessed. At first, the V parameter and $\%bias$ were estimated. Figure 2 shows the average \bar{V} estimates as a function of the true V , evaluated for three different λ values ($\lambda = 2, 10, 18$) with three different sample size n ($n=500,1000,5000$). The best linear fitting was evaluated using the WLS fitting to eq. (5); each estimation was repeated $N = 50$ and $N = 100$ times. On each plot the dotted line represents the true $\bar{V} = V$ regression line; it was used as a reference. The error bars represent the se . Figure 3 shows the $\%bias$ evaluated on the data of Figure 2. The accuracy of the λ estimates was evaluated and the results are shown in Figures 4 and 5. Figure 4 shows the plot of the average $\bar{\lambda}$ estimates as a function of the true λ at three V values ($V = 1.3, 1.0, 0.7$); the error bars represent the se . The relevant $\%bias$ is shown in Figure 5. Figure 6 shows the results of two methods used to characterize the deviation from the Poisson distribution: 1) the dispersion index (D) as a function of $\hat{\lambda}$ and \hat{V} at three lambda values ($\lambda = 2, 10, 18$); 2) the \hat{v} estimate as a function of the true V . As expected, the intersection between D and \hat{v} curves is at about $V = 1$ (Poisson process). It confirms that the method based on the moments almost agrees with the method based on eq. (5). Unfortunately, the first one does not supply any information on the statistical model supporting the data, that is one of the goals of this paper.

3.2 Fitting the CMP distribution to phantom data

In this section, we illustrate the applicability of the CMP-WLS distribution to fit PET sinogram data derived from the acquisition of an experimental phantom. PET data consist of six volumetric data; the first one includes raw-data (not-corrected), the remaining ones contain data corrected for physical and geometric artifacts. For each volumetric data the procedure described in section 2.2 was applied, to realize the ROIs with an adequate sample size n , and to use them in the regression analysis. Each ROI is constituted by $n = 315 * nl = 945$ data points, with $nl = 3$. Firstly, a CMP plot is built for each ROI using eq. (5), where x is the measured phantom activity. Then, a straight line is fitted to the data by using WLS method, after defining the amplitude range of $\log(p(x-1)/p(x))$ vs. $\log(x)$ with an almost constant spread, according to the simulation procedure. The dispersion \hat{v} and the centring $\hat{\lambda}$ parameters were derived from the linear fitting under the assumption of CMP

distribution. Such procedure was applied to 12 ($nv/nl=36/3$) independent ROIs. For each volumetric data, the mean value \bar{v} over the 12 ROIs was evaluated. This procedure was repeated on each volumetric data as a function of the phantom activity, from about 2.3 to 5 KBq/ml. Figure 7a shows the trend of \hat{v} estimates along with the activity. The centring parameter $\hat{\lambda}$ estimate versus the activity was analyzed and plotted in Figure 7b. The figure 7c shows the dispersion index D vs. the activity.

4. Discussion

In the current work, we discussed the use of the CMP distribution model to assess the statistics of PET data. Parameters estimation based on the CMP model is a well-recognized approach to quantitatively characterize deviation from Poisson distribution in many fields [24-33]. However, to our knowledge, the use of CMP distribution model for characterizing PET data statistics remains unexplored. A different approach aimed to characterize the statistical behavior of PET measurements, pre-corrected for attenuation and accidental coincidences, was proposed in [38]. Authors assumed the formation of over-dispersed data as due to corrections and modeled them using the Negative Binomial (NB) distribution. The weakness of the NB distribution is that it is not suitable to describe under-dispersed data, which instead can be handled with the CMP distribution. In [9] authors evaluated the statistical properties of the reconstructed PET data, according to row-action maximum likelihood algorithm (RAMLA) and filtered back-projection (FBP) reconstruction techniques. They stated that PET data, after correction and reconstruction, are no longer Poisson distributed; moreover, they characterized the noise and the *std/mean* by using several statistical distributions including gamma, normal, and NB.

The method herein discussed, that is based on the CMP-WLS as described in the sections 2.1.1 to 2.1.3, appears simple and computationally efficient. Other methods have been proposed in the literature, based on maximum likelihood estimation or the estimation of the posterior probability of the parameters of the CMP distribution [33,39]. Undoubtedly such methods are more accurate and robust, but the relevant equations cannot be solved analytically, thus introducing a high and unacceptable level of complexity in the analysis of PET measurements. As far as the homoscedasticity assumption is verified, the proposed method allowed us to identify the interval where the spread of the dependent variable is almost constant. Such a simple approach appears

adequate to produce consistent estimates of the dispersion parameter $\hat{\nu}$ at typical dose used in the measured PET data.

We used Monte Carlo simulation of CMP distributed data to evaluate the goodness of the dispersion ν and centring λ parameters estimation, according to the CMP-based method. The average estimate of the $\bar{\nu}$ parameter as function of the true ν , shown in Figure 2, confirms a fairly linear relationship; as the sample size n increases, the linear relationship is getting closer to the reference line; the *se* reduces as n increases. The number of averaging N is less effective with respect to n . The *%bias* shown in Figure 3 decreases as n increases; the increase in N is less effective. By rejecting the lowest sample size ($n=500$) because too noisy, the points on the left tail experience a large *%bias*, which resolves in any case in almost a 10% error as ν increases. So the bias is worse in over- with respect to under-dispersed data. In summary, the *%bias* decreases below a 10% error as ν increases in the worst case ($\lambda = 18$) and it further decrease with decreasing λ . This last finding allows us to stress that at any n and N values, the ν estimates are highly accurate in low emission counting, the most critical range in clinical PET measurements [17,18,19,20]. As far as the centring parameter λ is concerned, as documented in Figure 4, the average $\bar{\lambda}$ estimate maintains an almost linear relationship with respect to the true λ , except for points on the right tail, which correspond to high λ values estimates. Deviation from linear regression is more marked at low N ($N=500$) and $n(n=50)$ values. As expected, the *se* reduces as n and N increase. The *%bias*, as it is shown in Figure 5, decreases as n and N increase; at any n and N , it drastically decreases with decreasing λ . In the worst case (high λ), to preserve a maximum 20% error, the sample size n must be, at least, equal to 1000. The estimation accuracy improves consequently increasing the sample size. Such evidence was used in choosing the minimum number of samples when estimating λ on measured PET data. Previous results point out a better estimate of the λ parameter in low emission count at any ν ; for higher emission, a better accuracy is a direct consequence of the sample size. Figure 6 summarizes the overall behavior of the ν and λ estimates regarding the moments of the CMP distribution. As expected, the intersection between under- and over-dispersion curves is around $\nu = 1$ (Poisson process), where D is about one; it spreads as ν deviates from Poisson. It highlights the sensitivity of the method based on eq. (5) to quantitatively describe the deviation from Poisson on simulated data. Such trend was observed at any estimated λ value. In

summary, simulation results show that deviation from Poisson statistics in raw data and data after corrections are detectable with a CMP-based method.

We tested the method on experimental PET data confirming that the CMP model can assess the statistics of raw and corrected PET data, provided that data selection is performed according to simulation results. First of all, the data size n should be at least equal to 1000 to obtain reliable statistical parameters estimation; this is the reason we used $nl=3$ lines in the experimental phantom data analysis, obtaining data size $n=945$. Data averaging is preferred to improve the estimates; for this reason, the phantom data was obtained as the mean of $N=nl=12$ ROIs data. Accordingly, the raw and corrected phantom data were processed and the results are shown in Figure 7. The error bars ($\pm se$) shown in figure 7a confirm an almost homoscedasticity in the amplitude range selected for the analysis and a low spread in the estimates. The $\hat{\nu}$ estimates do not seem to follow a Poisson model even in raw-data before correction; such evidence is still not sufficiently discussed in the literature, mainly focused on the assumption of the Poisson model. The direct consequence of corrections on sinogram data appears in the presence of over- ($\nu < 1$) and under- ($\nu > 1$) dispersed data. Such circumstances occur at any activity value; as the activity increases, the consequent decrease in $\hat{\nu}$ estimates underlines a further deviation from Poisson statistics, which appears both in raw data and in data after random and deadtime corrections. The last evidence needs to be answered. Data after scatter correction appear closer to a Poisson distribution. To fully characterize PET data in terms of the CMP model, the $\hat{\lambda}$ estimate versus the activity was analyzed and plotted in Figure 7b. Unsurprisingly, the trend in $\hat{\lambda}$ estimates increases with the emission activity for any data (i.e. raw data and data after correction). The figure 7c, that is the consequence of the application of eq. (3) and eq. (4), offers an easier interpretation of data in terms of the moments of the CMP distribution, but it does not supply evidence on the statistical model.

5. Conclusions

The CMP distribution allows more flexibility in modelling PET data with respect to the Poisson distribution. Thanks to its flexibility in fitting over- and under-dispersed dataset, supported by the shape parameter ν , the CMP distribution appears a formidable tool in tracing the statistical behavior

of PET data during the several procedural corrections involved in PET reconstruction process. This paper documented the performance of CMP distribution for PET dataset using both simulation and measured data. The results of this study demonstrated that the CMP distribution **could** handle over-, equi-, and under-dispersed data with an adequate accuracy. The method herein discussed seems to have higher efficiency in low counting emission data, as it happens on dynamic PET data, or in ‘cold spot’ images, or in **low** dose injection studies. Measured PET data confirm deviations from Poisson model both in raw and in corrected data, stressing the need for a flexible statistical model able to describe measured PET data.

Although several research objectives have been achieved, further issues exploiting the CMP model to include image reconstruction and quantitative analysis need to be investigated in future.

In conclusion, this research may open new insight in the interpretation of clinical data, provided that statistical models able to account for deviations from the Poisson model are taken into consideration.

Figure captions

Figure 1. typical trends of $\log(p(x-1)/p(x))$ from eq. (5) vs $\log(x)$ for four different centring parameter λ ($\lambda = 5, 10, 15, 20$) and dispersion parameter $V = 1$.

Figure 2. estimated dispersion parameter \bar{V} vs true value V (in log scale), for three different values of the centring parameter λ ($\lambda = 2, 10, 18$); sample size $n= 500$ in the first line, $n=1000$ in the middle line, $n=5000$ in the last line; $N = 50$ in the left column, $N= 100$ in the right column. The dotted reference line is also shown, representing the true $\bar{V} = V$ regression line

Figure 3. %bias on estimating the dispersion parameter values \bar{V} vs true values V (in log scale), for three different values of the centring parameter λ ($\lambda = 2, 10, 18$); sample size $n= 500$ in the first line, $n=1000$ in the middle line, $n=5000$ in the last line; $N = 50$ in the left column, $N= 100$ in the right column

Figure 4. estimated centring parameter $\bar{\lambda}$ vs true values λ at three values of the dispersion parameter V ($V = 1.3, 1, 0.7$); sample size $n= 500$ in the first line, $n=1000$ in the middle line, $n=5000$ in the last line; $N = 50$ in the left column, $N= 100$ in the right column. The dotted reference line is also shown, representing the true $\bar{\lambda} = \lambda$ regression line

Figure 5. %bias on estimating the centring parameter values $\bar{\lambda}$ estimates vs true values λ at three values of the dispersion parameter V ($V = 1.3, 1.0, 0.7$); sample size $n= 500$ in the first line, $n=1000$ in the middle line, $n=5000$ in the last line; $N = 50$ in the left column, $N= 100$ in the right column

Figure 6. dispersion index D (continuous lines) and dispersion parameter \hat{V} (dotted lines) at three values of the centring parameter λ ($\lambda =2, 10, 18$) as a function of the true values V .

Figure 7. experimental results; a: dispersion parameter \hat{V} estimates vs activity; b: centring parameter $\hat{\lambda}$ estimates vs activity; c: dispersion index D vs activity.

Summary

Positron emission tomography (PET) in medicine exploits the properties of positron-emitting **unstable** nuclei. The pairs of γ - rays emitted after annihilation are revealed by coincidence detectors and stored as projections in a sinogram. It is well known that radioactive decay follows a Poisson distribution; however, deviation from Poisson statistics occurs on PET projection data **before** reconstruction due to physical effects, measurement errors, correction of deadtime, scatter and random coincidences.

When deviation from Poisson model is notable, Poisson assumption cannot be applied anymore. Such trade-off could be suitably steered by the knowledge of the true statistical model of sinogram data. However, the degree of deviation from Poisson model **remains** unknown for the lack of a reliable method that is able to quantify such deviations. A model that describes the statistical behaviour of measured and corrected PET data can aid in understanding the statistical nature of the data; it could lead to a more appropriate correction, reducing bias and variability and it is a prerequisite to **develop** efficient reconstruction and processing methods and to reduce noise.

In the present paper, the deviation from Poisson statistics in PET data is described by the Conway-Maxwell-Poisson (CMP) distribution model, which is characterized by the centring parameter and the dispersion **parameter**, the latter quantifying the deviation from a Poisson distribution model. In fact, the parameter allows to quantify over-dispersion (<1), equi-dispersion ($=1$), or under-dispersion (>1) of data.

A simple and efficient method for parameters estimation has been introduced in the present work and assessed using Monte Carlo simulation in a wide range of activity values.

The CMP distribution model herein proposed for PET data processing, has already been studied extensively in other disciplines such as Biology, Transportation, Marketing, eCommerce; to our knowledge, the use of CMP distribution model for characterizing PET data statistics remains unexplored. So the extension of the CMP model to analyze PET sinogram data is novel.

The proposed method has been applied both on simulated and experimental PET phantom data. The results of this study demonstrated that the CMP distribution **could** handle over-, equi-, and under-dispersed data with an adequate accuracy allowing to consider it a formidable tool in tracing the statistical behaviour of PET data during the several procedural corrections involved in PET reconstruction process.

It may be usefully implemented in image reconstruction algorithms and quantitative PET data analysis, especially in low counting emission data, as in dynamic PET data, where the method demonstrated the best accuracy.

References

-
- [1] M.Rahimi-Gorji, T.B.Gorji, M.Gorji-Bandpy. Details of regional particle deposition and airflow structures in a realistic model of human tracheobronchial airways: two-phase flow simulation, *Computers in Biology and Medicine*. 74:1 (2016) 1-17.
 - [2] M. Rahimi-Gorji, O. Pourmehran, M. Gorji-Bandpy, T.B. Gorji. CFD simulation of airflow behavior and particle transport and deposition in different breathing conditions through the realistic model of human airways, *Journal of Molecular Liquids*, 209 (2015) 121-133.
 - [3] F. Gibiino, F. Positano, V. Wiesinger, F. Giovannetti, G. Landini, L. Santarelli, MF(2013). Structured errors in reconstruction methods for Non-Cartesian MR data. *Computers in Biology and Medicine*, 43:12 (2013) 2256-2262.
 - [4] F. Gibiino, V. Positano, L. Landini, M.F. Santarelli. Regularization techniques on least squares non-uniform fast Fourier transform. *International Journal for Numerical Methods in Biomedical Engineering*. 29:5 (2013) 561 - 573
 - [5] Y.Vardi, L. A.Shepp, L.Kaufman. A Statistical Model for Positron Emission Tomography”, *Journal of the American Statistical Association*. 80:389 (1985) 8-20.
 - [6] B.Guérinand G.E. Fakhri. Realistic PET Monte Carlo Simulation With Pixelated Block Detectors, Light Sharing, Random Coincidences and Dead-Time Modeling. *IEEE Trans Nucl Sci*. 55:3 (2008) 942–952 .
 - [7] S.R.Cherry, J. A.Sorenson, M. E.Phelps. *Physics in nuclear medicine*. Elsevier Saunders Eds (2012).
 - [8] J.Qi and R. H.Huesman. Effect of Errors in the System Matrix on MAP Image Reconstruction. *Phys. Med Biol*. 50:14 (2005) 3297–3312.
 - [9] T.Teymurazyan, H.S.Riauka, D. R.Jans. Properties of Noise in Positron Emission Tomography Images Reconstructed with Filtered-Backprojection and Row-Action Maximum Likelihood Algorithm. *J Digit Imaging*. 26 (2013) 447–456.
 - [10] H. Zaidi (ed). *Quantitative Analysis in Nuclear Medicine imaging*. New York: Springer (2006).
 - [11] M.Dawood, X.Jiang and K.Schäfers (ed). *Correction Techniques in Emission Tomography*, Boca Raton, CRC Press, Series in Medical Physics and Biomedical Engineering (2012).

-
- [12] D.Brasse, P.E.Kinahan, C.Lartizien, C.Comtat, M.Casey and C.Michel. Correction Methods for Random Coincidences in Fully 3D Whole-Body PET: Impact on Data and Image Quality. *J Nucl Med* 46 (2005) 859–867.
- [13] M.L. Larsenand, A.B.Kostinski. Simple dead-time corrections for discrete time series of non-Poisson data. *Meas. Sci. Technol.* 20 (2009) 1-10.
- [14] F.Luisier, T.Blu, and M.Unser. Image Denoising in Mixed Poisson–Gaussian Noise. *IEEE Transactions On Image Processing*, 20:3 (2011) 696-708.
- [15] S.Ahn, and J. A.Fessler.Emission Image Reconstruction for Randoms-Precorrected PET Allowing Negative Sinogram Values, *IEEE Transactions on Medical Imaging*, 23:5 (2004) 591-601.
- [16] M. E.Casey. Point Spread Function Reconstruction in PET, Siemens Medical Solutions USA (2007).
- [17] Y.Jian, B.Planeta, R E.Carson. Evaluation of bias and variance in low-count OSEM list mode reconstruction. *Phys. Med. Biol.* 60 (2015) 15–29.
- [18] A. Reilhac, S. Tomeï, I. Buvat, C. Michel, F.Keheren, N.Costes. Simulation-based evaluation of OSEM iterative reconstruction methods in dynamic brain PET studies. *Neuroimage.* 39:1 (2008) 359-68.
- [19] M. D.Walker, M. C.Asselin, P. J.Julyan, M.Feldmann, P. S.Talbot, T.Jones, J. C.Matthews. Bias in iterative reconstruction of low-statistics PET data: benefits of a resolution model. *Phys. Med. Biol.* 56 (2011) 931–949.
- [20] G.L. Zeng. The ML-EM Algorithm is Not Optimal for Poisson Noise. *IEEE Transactions on Nuclear Science.* 62:5 (2015) 2096 - 2101.
- [21] J. D. Castillo and M. Pérez-Casany, Overdispersed and underdispersed Poisson generalizations, *Journal of statistical Planning and Inference*, 134 (2005), 486-500.
- [22] Krestyannikov E., Ruotsalainen U. (2004). Quantitatively accurate data recovery from attenuation-corrected sinogram using filtering of sinusoidal trajectory signals. *IEEE Nuclear Science Symposium*, 5, 3195-3199.
- [23] M.F. Santarelli, V. Positano, L. Landini. Measured PET Data Characterization with the Negative Binomial Distribution Model. *Journal of Medical and Biological Engineering.* (2016) In Press
- [24] R.A. Francis , S.R. Geedipally, S.D. Guikema, S.S. Dhavala, D. Lord , S. La Rocca . Characterizing the performance of the Conway-Maxwell Poisson generalized linear model. *Risk Anal.* 32:1 (2012) 167-83.
- [25] K.F. Sellers, G. Shmueli. Data Dispersion: Now You See It... Now You Don't. *Communications in Statistics - Theory and Methods.* 42:17 (2013) 3134-3147.
- [26] K.F. Sellers, S.Borle, G.Shmueli. The COM-Poisson Model for Count Data: A Survey of Methods and Applications. *Applied stochastic models in business and industry.* 28:2 (2012) 104-116.

-
- [27] M.Asalam, M.Azamb and C.H. Junc. A Control Chart for COM–Poisson Distribution Using Resampling and Exponentially Weighted Moving Average. *Qual. Reliab. Engng. Int.* 32 (2016) 727–735.
- [28] P.Sur, G.Shmueli, S.Bose & P.Dubey. Modeling Bimodal Discrete Data Using Conway-Maxwell-Poisson Mixture Models. *Journal of Business & Economic Statistics*, 33:3 (2015) 352-365.
- [29] M.S.Ridout, P.Besbeas. An empirical model for underdispersed count data. *Statistical Modelling*, 4 (2004) 77–89.
- [30] D.Lord, S.R.Geedipally, S.D.Gaikema. Extension of the application of Conway–Maxwell–Poisson models: analyzing traffic crash data exhibiting underdispersion. *Risk Analysis* . 30 (2010) 1268–1276.
- [31] P.Boatwright, S.Borle, J.B. Kadane. A model of the joint distribution of purchase quantity and timing. *Journal of the American Statistical Association*. 98 (2003) 564–72.
- [32] S. Singh, S. Borle, D. Jain. A generalized framework for estimating customer lifetime value when customer lifetimes are not observed. *Quantitative Marketing and Economics*. 7 (2009) 181–205.
- [33] G. Shmueli, T. P. Minka, J. B. Kadane, S. Borle, and P. Boatwright. A Useful Distribution for Fitting Discrete Data: Revival of the Conway-Maxwell-Poisson Distribution. *Applied Statistics*, 54 (2005) 127-142.
- [34] A. Papoulis, *Probability, Random Variables, and Stochastic Processes* 4th edn. McGraw-Hill (2002).
- [35] D. R. Cox, P. A. W. Lewis. *The Statistical Analysis of Series of Events*. London: Methuen(1966).
- [36] Fano U. (1947). Ionization Yield of Radiations. II. The Fluctuations of the Number of Ions. *Phys. Rev.* 72:26.
- [37] R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria (2013) URL <http://www.R-project.org/>.
- [38] E. Krestyannikov, U. Ruotsalainen. Quantitatively accurate data recovery from attenuation-corrected sinogram using filtering of sinusoidal trajectory signals. *IEEE Nuclear Science Symposium*, 5 (2004) 3195-3199.
- [39] F. K. Sellers, S. Borle, and G. Shmueli. The COM-Poisson model for count data: a survey of methods and applications. *Appl. Stochastic Models Bus. Ind.*, 28 (2012) 104–116.

A CONWAY-MAXWELL-POISSON (CMP) MODEL TO ADDRESS DATA DISPERSION ON POSITRON EMISSION TOMOGRAPHY

Maria Filomena Santarelli^{*a,b}, Daniele Della Latta^b, Michele Scipioni^c, Vincenzo Positano^b, Luigi Landini^{c,b,a}

a. Institute of Clinical Physiology, National Research Council, via Moruzzi 1, 56124 Pisa, Italy;

b. Fondazione CNR-Regione Toscana "G. Monasterio", via Moruzzi, 1, 56124 Pisa, Italy;

c. Department. of Information Engineering: EIT, University of Pisa, via Diotisalvi 2, 56126 Pisa, Italy.

* Correspondence to: Institute of Clinical Physiology, National Research Council, via Moruzzi 1, 56124 Pisa, Italy Tel.: +39 050 3152614; fax: +39 050 3152166; e-mail: santarel@ifc.cnr.it (M.F. Santarelli)

ABSTRACT

Positron emission tomography (PET) in medicine exploits the properties of positron-emitting unstable nuclei. The pairs of γ - rays emitted after annihilation are revealed by coincidence detectors and stored as projections in a sinogram. It is well known that radioactive decay follows a Poisson distribution; however, deviation from Poisson statistics occurs on PET projection data prior to reconstruction due to physical effects, measurement errors, correction of deadtime, scatter, and random coincidences. A model that describes the statistical behaviour of measured and corrected PET data can aid in understanding the statistical nature of the data: it is a prerequisite to develop efficient reconstruction and processing methods and to reduce noise.

The deviation from Poisson statistics in PET data could be described by the Conway-Maxwell-Poisson (CMP) distribution model, which is characterized by the centring parameter λ and the dispersion parameter ν , the latter quantifying the deviation from a Poisson distribution model. In particular, the parameter ν allows quantifying over-dispersion ($\nu < 1$) or under-dispersion ($\nu > 1$) of data. A simple and efficient method for λ and ν parameters estimation is introduced and assessed using Monte Carlo simulation for a wide range of activity values.

The application of the method to simulated and experimental PET phantom data demonstrated that the CMP distribution parameters could detect deviation from the Poisson distribution both in raw and corrected PET data. It may be usefully implemented in image reconstruction algorithms and quantitative PET data analysis, especially in low counting emission data, as in dynamic PET data, where the method demonstrated the best accuracy.

KEYWORD: positron emission tomography (PET), Conway-Maxwell-Poisson (CMP) distribution, maximum likelihood (ML) estimation, sinograms, Poisson statistic deviation.

Nomenclature	
CMP	Comway-Maxwell-Poisson
ML	maximum likelihood
ML-EM	maximum likelihood expectation maximization
NB	negative binomial
PET	positron emission tomography
WLS	weighted least square
Se	standard error
$\%bias$	bias index in percent
$E[X]$	mean of the CMP process
$Var[X]$	variance of the CMP process
D	dispersion index
n	sample size
N	number of repetitions of the same experiment
Nv	number of vectors, i.e. central line of the sinograms, used in the analysis of experimental phantom data
Nl	number of sequenced vectors, used in the analysis of experimental phantom data
λ	centring parameter in the CMP distribution; it is related to the observations mean
ν	shape parameter in the CMP distribution; it is related to data over-dispersion and under-dispersion
$\hat{\lambda}$	estimated centring parameter over the sample size n
$\bar{\lambda}$	mean of the $\hat{\lambda}$ value, over N repetitions
$\hat{\nu}$	estimated dispersion (shape) parameter over the sample size n
$\bar{\nu}$	mean of the $\hat{\nu}$ value, over N repetitions
θ	dummy variable, stands for ν and λ
$\hat{\theta}_i$	estimated value of θ at each repetition

1. Introduction

Today numerical and computational methods are becoming increasingly frequent in biology and medicine to gain new insights into the emerging research trends [1-4]. Among those benefiting from the numerical and computational methods, medical imaging plays an integral role, especially in the image formation process. In particular, in PET imaging such methods are crucial to grasp both the biological and the technological aspects of image acquisition and reconstruction processes. Actually, in positron emission tomography (PET) a tracer containing a radioactive isotope is introduced into the body, and it emits positrons while being involved in metabolic processes. Due to annihilation between a positron and an electron, a pair of γ -rays is emitted in opposite directions along a line of response (LOR) at each spatial coordinate; the γ -ray pair is revealed by a coincidence detectors pair placed in an array of detector units surrounding the body. Emission of γ -rays follows a Poisson law [5]. In this paper, numerical and computational methods are used to investigate deviation from Poisson statistics occurring in PET projection data before (or during) reconstruction.

A known source of deviation from Poisson model is related to errors introduced in correcting some unwanted physical and acquisition effects, such as scattering errors, accidental errors, positron range, deadtime, detectors non-collinearity, variation in detection-pair sensitivity, and others. [6-9]. Scattered coincidences are due to one or multiple Compton scatterings of the detected photons; as a result, wrong LORs are obtained, which cause a reduced contrast and an overestimation of the tracer activity [10]. Random (or false) coincidences occur when photons detected within the same time window are not generated from the same annihilation event. Random events cause a reduction of image contrast and a bias in the measured activity due to an increased level of background activity [11]. Both scattering and random coincidences are not wanted events, which need a correction for obtaining exact emission data; however, such corrections cause deviation from Poisson statistics [12]. A further source of deviation from Poisson law is the deadtime, i.e. the instrumental insensitivity to detectable events due to electronic or mechanical reset time. Especially in high counting rate, some events can be missed, and the detected rate is lower than the true one [13]. Reduction in the mean value is a potential cause of the deviation from the Poisson model. The detectors non-collinearity and the variation in detection-pair sensitivity are two scanner-related factors that cause an incorrect emission evaluation, and need correction [11]. The sensitivity of the LOR about the mean is affected both by the geometry of the detectors and the LOR position; hence,

a geometric correction is needed to get a correct detection. Besides such geometric effects, the block detectors themselves vary in efficiency, as the photomultiplier tubes gains are not uniform (and may change in time), and the crystals are not all identical; then a normalization operation is required. In commercial PET scanners, in addition to the stochastic nature of the photon-counting, the intrinsic thermal and electronic noise should be accounted for. Such sources of artefacts are usually described by a Poisson model corrupted with additive Gaussian noise [14]. When deviation from Poisson model is notable, Poisson assumption cannot be applied anymore. Such trade-off could be suitably steered by the knowledge of the true statistical model of sinogram data. Unfortunately, the degree of deviation from Poisson model remains unknown for the lack of a reliable method that can quantify such deviations.

Using a fully 3D maximum likelihood expectation maximization (ML-EM) reconstruction algorithm, two correction methods are currently used. In one method measured data are pre-corrected before applying the reconstruction algorithm. We know that pre-correction for random and scatter coincidences may result in deviation from Poisson assumption, depending on the entity of correction. It compromises the estimation of the emission density based on ML reconstruction [15]. The second correction method incorporates correction models into the iterative algorithm, thus preserving the Poisson nature of the data [16]. Such solution is not immune to errors due to lack of knowledge about the true statistics of sinogram data. Moreover, such approach produces bias and higher variability in applications where images are reconstructed from a relatively small number of counts [17-20]. This issue is an important aspect to take into account, especially for cold regions and/or dynamic PET data reconstruction; actually, bias and variability of reconstructed data can lead to an incorrect activity evaluation and inexact kinetic parameters estimate. Without any doubts, an accurate evaluation of the deviation of the measured PET data from Poisson statistics could lead to a more appropriate correction, reducing bias and variability.

In the literature, a few classes of discrete probability models were proposed and studied as alternatives to Poisson model in measured data. Sometimes, the measured data describe an over-dispersion or an under-dispersion as a deviation from the Poisson model [21]. In PET data analysis, the Negative Binomial (NB) distribution model has been proposed in monitoring the over-dispersion on both uncorrected or corrected (random and scatter coincidence correction) sinograms. [22,23]. In such papers, the authors highlighted the importance of an accurate modelling of PET data to develop efficient reconstruction algorithms. To account for over and under-dispersion

simultaneously, the Conway-Maxwell-Poisson (CMP) distribution model can be suggested. It is worth to note that the CMP distribution model has been studied extensively [24-28]. Applications of the CMP distribution model range from Biology [29] to Transportation [30], Marketing, and eCommerce [31,32]. However, to our knowledge, the use of CMP distribution model to describe PET data statistics remains unexplored. So, the application of the CMP model to the analysis of PET sinogram data is novel. The relevant benefits are twofold. Firstly, the CMP model allows to fully characterize the statistical behavior of PET sinogram data, by accounting for both uncorrected and corrected data, thus making possible the inclusion in the correction model of random, deadtime, normalization, geometric, and scatter coincidence effects. Secondly, the proposed approach may prelude to the application of the CMP model in the wide field of high energy medical imaging, in the situation where deviations from the conventionally used Poisson distribution model may happen.

This paper describes a method that uses the CMP model to assess the statistics of PET projections, collected into sinograms, in a wide range of emission rates. We focused on uncorrected and corrected (for random, deadtime, normalization, geometric, and scatter coincidence effects) sinograms, due to their importance in the quantitative analysis of PET data. A method that accurately defines the statistics of PET measured data, also in the conditions of deviations from Poisson statistics, could have important implications for the quantitative evaluation of PET data.

2. Materials and Methods

2.1 The Conway-Maxwell-Poisson (CMP) distribution model

2.1.1 The CMP probability function and its moments

The importance of CMP distribution model is due to its ability to model count data with different degrees of over- and under-dispersion. This distribution is a two-parameters extension of the Poisson distribution. Its probability density function is [24]:

$$P(X = x; \lambda, \nu) = \frac{\lambda^x}{(x!)^\nu} [Z(\lambda, \nu)]^{-1}, \quad \lambda > 0, \nu \geq 0, x = 0, 1, \dots$$

(1)

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}$$

where x denotes the amplitude data, λ is a centring parameter related to the observations mean, ν is the shape parameter of the CMP distribution, and Z is a normalizing constant.

The ν parameter outlines deviations from Poisson distribution: $\nu > 1$ corresponds to under-dispersed data, $\nu < 1$ to over-dispersed data, and $\nu = 1$ to equi-dispersed (Poisson) data. As $\nu = 1$ the mean value becomes equal to λ (the Poisson parameter).

Previous relationship (1) allows for the following representation as a non-linear decrease in ratio of successive probabilities [33]:

$$\frac{P(X = x-1)}{P(X = x)} = \frac{x^\nu}{\lambda} \quad (2)$$

Such formulation enables us to use the weighted least squares (WLS) method to estimate the CMP parameters. The λ and ν parameters can be used to derive an approximation of the mean, $E[X]$, and the variance, $\text{var}[X]$, of the CMP process according to the following relationships [33]:

$$E[X] = \lambda^{1/\nu} + \frac{1}{2\nu} - \frac{1}{2} \quad (3)$$

$$\text{var}[X] = \frac{1}{\nu} \lambda^{1/\nu} \quad (4)$$

It follows that, for $\nu = 1$, $E[X] = \lambda$ and $\text{var}[X] = \lambda$, which corresponds to a Poisson process.

2.1.2 CMP parameters estimation method

The λ and ν parameters can be estimated using the approach described in [33]. The approach is based on eq. (2), relating the ratio of successive CMP probabilities to a non-linear relationship with

the λ and ν parameters. The following regression equation is found, taking the logarithm of both sides of eq.(2) [26]:

$$\log(p(x-1)/ p(x)) = -\log(\lambda) + \nu \log(x) \quad (5)$$

where $p(x)$ denotes $P(X=x)$. In actual situations, the ratio of probabilities on the left side of eq.(5) can be substituted with the relative frequency of occurrence of $(x-1)$ and x . The amplitude x is divided into a number of intervals, or classes, and for each class, the frequency of occurrence is evaluated. The number of classes is chosen in order to avoid zero counts in every class.

The estimation process can be split into two steps: 1) the term on the left side is plotted against the $\log(x)$ for every value of x ; 2) the best linear fitting is evaluated according to the WLS method to extract the CMP parameters. In the ideal CMP model, the data lie in a straight line, with angular coefficient ν and intercept $\log(\lambda)$. In the Poisson model ($\lambda=1, \nu=1$) the intercept would be equal to zero and the angular coefficient would be equal to one.

The percent-bias (*%bias*) and standard error (*se*) were used as performance indices of λ and ν parameters estimator. For each parameter, the *%bias* is defined as the difference between the expected and the true value. When *%bias* is zero, the estimator is unbiased. The *%bias* index is defined as follows[34]:

$$\%bias = \frac{1}{N} \sum_{i=1}^N \hat{\theta}_i - \theta}{\theta} \times 100 \quad (6)$$

where θ is a dummy variable and it stands for ν and λ ; the mean computation is performed over N repetitions of the same experiment; $\hat{\theta}_i$ is estimated at each repetition over a sample of size n and θ is the true value.

The 'se' index is defined as the ratio between the standard deviation (*std*) of $\hat{\theta}_i$ ($i=1,..N$) estimates and the square root of the sample size[35]: $se = std/\sqrt{N}$.

The dispersion index, or Fano factor, D is defined as follows [35,36]:

$$D = \text{var}[X | \hat{\lambda}, \hat{\nu}] / E[X | \hat{\lambda}, \hat{\nu}] \quad (7)$$

with $E[X | \hat{\lambda}, \hat{\nu}]$ and $\text{var}[X | \hat{\lambda}, \hat{\nu}]$ calculated according to eq. (3) and eq. (4). It is an index which gives information about the over- or under-dispersion of data. The Poisson process is characterized by $D = 1$.

2.1.3 Data homoscedasticity

The ordinary regression model herein used is based on the assumption of homoscedasticity of data amplitude distributed on the frequency of occurrence classes. Because, as demonstrated in [33], such assumption is violated in the CMP distributed data (the variance of the dependent variable is not constant), we expect implications on parameters estimation accuracy. Firstly, the CMP plot is constructed plotting the logarithm of the ratio of successive frequencies ($\log(p(x-1)/p(x))$ vs $\log(x)$) for various combinations of n , λ and ν . On such a plot we evaluated the regression parameters over the full amplitude range of $\log(x)$. After that, we progressively reduced such interval symmetrically with respect to the amplitude of the maximum frequency of occurrence (i.e. the highest $P(X=x)$). For each amplitude range, the *std* of the $\hat{\nu}$ and $\hat{\lambda}$ estimates was evaluated. The amplitude interval of $\log(x)$ with an almost constant *std*, at each estimated parameters, was assumed as the best compromise for homoscedasticity, and it was used in the following simulations. With the constraint defined above, the parameters were estimated by fitting a linear regression to eq.(5).

2.1.4 Characterization of the CMP parameters by simulation

To test the estimation accuracy of the CMP-WLS model, we simulated a dataset characterized by a CMP distribution with known regression parameters. Such parameters correspond to a range of mean and variance values typical of clinical PET data. We then estimated the *%bias* and the *se* of the estimates using Monte Carlo simulation.

The Monte Carlo simulation was implemented according to the following steps:

1. Generation of repeated CMP data, combining the following parameters: sample size, n , ranging from 500 to 5000; true centring parameter, λ , ranging from 1 to 20; ν value, ranging from 0.75 to 1.5. The excursion of the ν parameter allows simulating data with over-, equi- and under-dispersion encountered in PET sinogram data. In particular, the variance ranges from $0.1 * \lambda$ to $10 * \lambda$.

The **R** software [37] with compGLM package was used for data generation.

2. For each combination of n , λ and V , the following estimations were performed:
 - \hat{V} estimation over the sample size n for three λ values using eq. (5); each estimation was repeated $N = 50$ and 100 times;
 - $\hat{\lambda}$ estimation over the sample size n for three V values using eq. (5); each estimation was repeated $N = 50$ and 100 times;
 - evaluation of the mean \bar{v} over the \hat{v} values previously estimated;
 - evaluation of the mean $\bar{\lambda}$ over the $\hat{\lambda}$ values previously estimated;
 - evaluation of *%bias* and $\pm se$ of \hat{v} and $\hat{\lambda}$;

Parameters estimation and analysis were accomplished by using Matlab (The MathWorks, Inc., Natick, Massachusetts, United States) software.

2.2 Phantom data acquisition and analysis

A cylindrical phantom, with a radius of 12 cm and height of 15 cm, was filled with ^{18}F -FDG and water solution (34MBq of total activity at the beginning of the experiment). The phantom was put inside a PET/CT (General Electric Discovery) tomograph, exactly in the middle of the bore, with its long axis parallel to the bore's axis.

The experiment involved the acquisition of 47 slices of 2D PET data projections covering the phantom, for six different times, with a time interval of about 15-20 minutes. Each volumetric data was acquired in a total time of 300 sec. Due to the emission decay, the total activity at the beginning of each volumetric acquisition was reduced consequently; so, the total emission at the beginning of each time acquisition ranged from 34MBq to 15MBq.

For each time interval, the 2D sinograms consisted of 315x367 values for 47 slices; the following volumetric sinograms were stored:

- raw-data, without any correction, and:
 - o after correction for random counts
 - o after correction for deadtime

- after normalization
- after geometric correction
- after correction for scattering

The sequence of corrections was applied according to a protocol implemented in a commercial scanner.

On each sinogram, its central line was selected and organized in a vector of 315 points. For a homogeneous circular object, the central line of the sinogram is relevant to the sum of the phantom contributions situated along its diameter, at each angle. A total of 47 vectors were available for the analysis. From these, 11 vectors were removed (they are relevant to 6 slices on one side and 5 slices on the other side of the phantom), leaving $nv=36$ vectors for the analysis. The Region-Of-Interest (ROI) for the regression analysis was realized by sequencing more vectors to increase the estimation accuracy: the ROI is constituted of $n = 315*nl$ data points, where nl is the number of sequenced vectors. The optimal nl value was determined according to simulation results.

The estimation of the \hat{V} and $\hat{\lambda}$ parameters was then performed on the selected ROI. In particular, for each volumetric data, the estimated parameters were averaged over $N = nv/nl=36/nl$ independent ROIs, each one consisting of n data points covering the middle of the phantom; the averaging can be done thanks to the homogeneity of the phantom. Such procedure was repeated for each corrected and not-corrected sinogram, and for each emission dose.

3. Results

3.1 CMP parameters estimation on simulated data

This section deals with the assessment of the performance of the proposed CMP parameters estimation method, using Monte Carlo simulation. Firstly, the amplitude interval of $\log(x)$ with a constant spread on the estimated parameters was identified. In Figure 1 four typical trends (plus the error bars (*se*)) of the logarithm of the ratios of successive frequencies ($\log(p(x-1)/p(x))$) of eq. (5)) vs $\log(x)$, for four λ values ($\lambda = 5, 10, 15, 20$), $n=1000$, and $V = 1$, are shown. Such plots highlight the spread (i.e. the *se*) of the dependent variable as a function of $\log(x)$: at the extremes,

the spread is larger than in the middle of $\log(x)$. So, for each λ value, an interval of $\log(x)$ data with almost constant se was identified to be used in the following estimates. Then, the accuracy on V and λ estimation was assessed. At first, the V parameter and $\%bias$ were estimated. Figure 2 shows the average \bar{V} estimates as a function of the true V , evaluated for three different λ values ($\lambda = 2, 10, 18$) with three different sample size n ($n=500,1000,5000$). The best linear fitting was evaluated using the WLS fitting to eq. (5); each estimation was repeated $N = 50$ and $N = 100$ times. On each plot the dotted line represents the true $\bar{V} = V$ regression line; it was used as a reference. The error bars represent the se . Figure 3 shows the $\%bias$ evaluated on the data of Figure 2. The accuracy of the λ estimates was evaluated and the results are shown in Figures 4 and 5. Figure 4 shows the plot of the average $\bar{\lambda}$ estimates as a function of the true λ at three V values ($V = 1.3, 1.0, 0.7$); the error bars represent the se . The relevant $\%bias$ is shown in Figure 5. Figure 6 shows the results of two methods used to characterize the deviation from the Poisson distribution: 1) the dispersion index (D) as a function of $\hat{\lambda}$ and \hat{V} at three lambda values ($\lambda = 2, 10, 18$); 2) the \hat{V} estimate as a function of the true V . As expected, the intersection between D and \hat{V} curves is at about $V = 1$ (Poisson process). It confirms that the method based on the moments almost agrees with the method based on eq. (5). Unfortunately, the first one does not supply any information on the statistical model supporting the data, that is one of the goals of this paper.

3.2 Fitting the CMP distribution to phantom data

In this section, we illustrate the applicability of the CMP-WLS distribution to fit PET sinogram data derived from the acquisition of an experimental phantom. PET data consist of six volumetric data; the first one includes raw-data (not-corrected), the remaining ones contain data corrected for physical and geometric artifacts. For each volumetric data the procedure described in section 2.2 was applied, to realize the ROIs with an adequate sample size n , and to use them in the regression analysis. Each ROI is constituted by $n = 315 * nl = 945$ data points, with $nl = 3$. Firstly, a CMP plot is built for each ROI using eq. (5), where x is the measured phantom activity. Then, a straight line is fitted to the data by using WLS method, after defining the amplitude range of $\log(p(x-1)/p(x))$ vs. $\log(x)$ with an almost constant spread, according to the simulation procedure. The dispersion \hat{V} and

the centring $\hat{\lambda}$ parameters were derived from the linear fitting under the assumption of CMP distribution. Such procedure was applied to 12 ($nv/nl=36/3$) independent ROIs. For each volumetric data, the mean value \bar{v} over the 12 ROIs was evaluated. This procedure was repeated on each volumetric data as a function of the phantom activity, from about 2.3 to 5 KBq/ml. Figure 7a shows the trend of \hat{v} estimates along with the activity. The centring parameter $\hat{\lambda}$ estimate versus the activity was analyzed and plotted in Figure 7b. The figure 7c shows the dispersion index D vs. the activity.

4. Discussion

In the current work, we discussed the use of the CMP distribution model to assess the statistics of PET data. Parameters estimation based on the CMP model is a well-recognized approach to quantitatively characterize deviation from Poisson distribution in many fields [24-33]. However, to our knowledge, the use of CMP distribution model for characterizing PET data statistics remains unexplored. A different approach aimed to characterize the statistical behavior of PET measurements, pre-corrected for attenuation and accidental coincidences, was proposed in [38]. Authors assumed the formation of over-dispersed data as due to corrections and modeled them using the Negative Binomial (NB) distribution. The weakness of the NB distribution is that it is not suitable to describe under-dispersed data, which instead can be handled with the CMP distribution. In [9] authors evaluated the statistical properties of the reconstructed PET data, according to row-action maximum likelihood algorithm (RAMLA) and filtered back-projection (FBP) reconstruction techniques. They stated that PET data, after correction and reconstruction, are no longer Poisson distributed; moreover, they characterized the noise and the *std/mean* by using several statistical distributions including gamma, normal, and NB.

The method herein discussed, that is based on the CMP-WLS as described in the sections 2.1.1 to 2.1.3, appears simple and computationally efficient. Other methods have been proposed in the literature, based on maximum likelihood estimation or the estimation of the posterior probability of the parameters of the CMP distribution [33,39]. Undoubtedly such methods are more accurate and robust, but the relevant equations cannot be solved analytically, thus introducing a high and unacceptable level of complexity in the analysis of PET measurements. As far as the homoscedasticity assumption is verified, the proposed method allowed us to identify the interval

where the spread of the dependent variable is almost constant. Such a simple approach appears adequate to produce consistent estimates of the dispersion parameter $\hat{\nu}$ at typical dose used in the measured PET data.

We used Monte Carlo simulation of CMP distributed data to evaluate the goodness of the dispersion ν and centring λ parameters estimation, according to the CMP-based method. The average estimate of the $\bar{\nu}$ parameter as function of the true ν , shown in Figure 2, confirms a fairly linear relationship; as the sample size n increases, the linear relationship is getting closer to the reference line; the *se* reduces as n increases. The number of averaging N is less effective with respect to n . The *%bias* shown in Figure 3 decreases as n increases; the increase in N is less effective. By rejecting the lowest sample size ($n=500$) because too noisy, the points on the left tail experience a large *%bias*, which resolves in any case in almost a 10% error as ν increases. So the bias is worse in over- with respect to under-dispersed data. In summary, the *%bias* decreases below a 10% error as ν increases in the worst case ($\lambda = 18$) and it further decrease with decreasing λ . This last finding allows us to stress that at any n and N values, the ν estimates are highly accurate in low emission counting, the most critical range in clinical PET measurements [17,18,19,20]. As far as the centring parameter λ is concerned, as documented in Figure 4, the average $\bar{\lambda}$ estimate maintains an almost linear relationship with respect to the true λ , except for points on the right tail, which correspond to high λ values estimates. Deviation from linear regression is more marked at low N ($N=500$) and $n(n=50)$ values. As expected, the *se* reduces as n and N increase. The *%bias*, as it is shown in Figure 5, decreases as n and N increase; at any n and N , it drastically decreases with decreasing λ . In the worst case (high λ), to preserve a maximum 20% error, the sample size n must be, at least, equal to 1000. The estimation accuracy improves consequently increasing the sample size. Such evidence was used in choosing the minimum number of samples when estimating λ on measured PET data. Previous results point out a better estimate of the λ parameter in low emission count at any ν ; for higher emission, a better accuracy is a direct consequence of the sample size. Figure 6 summarizes the overall behavior of the ν and λ estimates regarding the moments of the CMP distribution. As expected, the intersection between under- and over-dispersion curves is around $\nu = 1$ (Poisson process), where D is about one; it spreads as ν deviates from Poisson. It highlights the sensitivity of the method based on eq. (5) to quantitatively describe the

deviation from Poisson on simulated data. Such trend was observed at any estimated $\hat{\lambda}$ value. In summary, simulation results show that deviation from Poisson statistics in raw data and data after corrections are detectable with a CMP-based method.

We tested the method on experimental PET data confirming that the CMP model can assess the statistics of raw and corrected PET data, provided that data selection is performed according to simulation results. First of all, the data size n should be at least equal to 1000 to obtain reliable statistical parameters estimation; this is the reason we used $n_l=3$ lines in the experimental phantom data analysis, obtaining data size $n=945$. Data averaging is preferred to improve the estimates; for this reason, the phantom data was obtained as the mean of $N=nl=12$ ROIs data. Accordingly, the raw and corrected phantom data were processed and the results are shown in Figure 7. The error bars ($\pm se$) shown in figure 7a confirm an almost homoscedasticity in the amplitude range selected for the analysis and a low spread in the estimates. The $\hat{\nu}$ estimates do not seem to follow a Poisson model even in raw-data before correction; such evidence is still not sufficiently discussed in the literature, mainly focused on the assumption of the Poisson model. The direct consequence of corrections on sinogram data appears in the presence of over- ($\nu < 1$) and under- ($\nu > 1$) dispersed data. Such circumstances occur at any activity value; as the activity increases, the consequent decrease in $\hat{\nu}$ estimates underlines a further deviation from Poisson statistics, which appears both in raw data and in data after random and deadtime corrections. The last evidence needs to be answered. Data after scatter correction appear closer to a Poisson distribution. To fully characterize PET data in terms of the CMP model, the $\hat{\lambda}$ estimate versus the activity was analyzed and plotted in Figure 7b. Unsurprisingly, the trend in $\hat{\lambda}$ estimates increases with the emission activity for any data (i.e. raw data and data after correction). The figure 7c, that is the consequence of the application of eq. (3) and eq. (4), offers an easier interpretation of data in terms of the moments of the CMP distribution, but it does not supply evidence on the statistical model.

5. Conclusions

The CMP distribution allows more flexibility in modelling PET data with respect to the Poisson distribution. Thanks to its flexibility in fitting over- and under-dispersed dataset, supported by the

shape parameter ν , the CMP distribution appears a formidable tool in tracing the statistical behavior of PET data during the several procedural corrections involved in PET reconstruction process. This paper documented the performance of CMP distribution for PET dataset using both simulation and measured data. The results of this study demonstrated that the CMP distribution could handle over-, equi-, and under-dispersed data with an adequate accuracy. The method herein discussed seems to have higher efficiency in low counting emission data, as it happens on dynamic PET data, or in 'cold spot' images, or in low dose injection studies. Measured PET data confirm deviations from Poisson model both in raw and in corrected data, stressing the need for a flexible statistical model able to describe measured PET data.

Although several research objectives have been achieved, further issues exploiting the CMP model to include image reconstruction and quantitative analysis need to be investigated in future.

In conclusion, this research may open new insight in the interpretation of clinical data, provided that statistical models able to account for deviations from the Poisson model are taken into consideration.

Figure captions

Figure 1. typical trends of $\log(p(x-1)/p(x))$ from eq. (5) vs $\log(x)$ for four different centring parameter λ ($\lambda = 5, 10, 15, 20$) and dispersion parameter $V = 1$.

Figure 2. estimated dispersion parameter \bar{V} vs true value V (in log scale), for three different values of the centring parameter λ ($\lambda = 2, 10, 18$); sample size $n= 500$ in the first line, $n=1000$ in the middle line, $n=5000$ in the last line; $N = 50$ in the left column, $N= 100$ in the right column. The dotted reference line is also shown, representing the true $\bar{V} = V$ regression line

Figure 3. %bias on estimating the dispersion parameter values \bar{V} vs true values V (in log scale), for three different values of the centring parameter λ ($\lambda = 2, 10, 18$); sample size $n= 500$ in the first line, $n=1000$ in the middle line, $n=5000$ in the last line; $N = 50$ in the left column, $N= 100$ in the right column

Figure 4. estimated centring parameter $\bar{\lambda}$ vs true values λ at three values of the dispersion parameter V ($V = 1.3, 1, 0.7$); sample size $n= 500$ in the first line, $n=1000$ in the middle line, $n=5000$ in the last line; $N = 50$ in the left column, $N= 100$ in the right column. The dotted reference line is also shown, representing the true $\bar{\lambda} = \lambda$ regression line

Figure 5. %bias on estimating the centring parameter values $\bar{\lambda}$ estimates vs true values λ at three values of the dispersion parameter V ($V = 1.3, 1.0, 0.7$); sample size $n= 500$ in the first line, $n=1000$ in the middle line, $n=5000$ in the last line; $N = 50$ in the left column, $N= 100$ in the right column

Figure 6. dispersion index D (continuous lines) and dispersion parameter \hat{V} (dotted lines) at three values of the centring parameter λ ($\lambda =2, 10, 18$) as a function of the true values V .

Figure 7. experimental results; a: dispersion parameter \hat{V} estimates vs activity; b: centring parameter $\hat{\lambda}$ estimates vs activity; c: dispersion index D vs activity.

Summary

Positron emission tomography (PET) in medicine exploits the properties of positron-emitting unstable nuclei. The pairs of γ - rays emitted after annihilation are revealed by coincidence detectors and stored as projections in a sinogram. It is well known that radioactive decay follows a Poisson distribution; however, deviation from Poisson statistics occurs on PET projection data before reconstruction due to physical effects, measurement errors, correction of deadtime, scatter and random coincidences.

When deviation from Poisson model is notable, Poisson assumption cannot be applied anymore. Such trade-off could be suitably steered by the knowledge of the true statistical model of sinogram data. However, the degree of deviation from Poisson model remains unknown for the lack of a reliable method that is able to quantify such deviations. A model that describes the statistical behaviour of measured and corrected PET data can aid in understanding the statistical nature of the data; it could lead to a more appropriate correction, reducing bias and variability and it is a prerequisite to develop efficient reconstruction and processing methods and to reduce noise.

In the present paper, the deviation from Poisson statistics in PET data is described by the Conway-Maxwell-Poisson (CMP) distribution model, which is characterized by the centring parameter and the dispersion parameter, the latter quantifying the deviation from a Poisson distribution model. In fact, the parameter α allows to quantify over-dispersion ($\alpha < 1$), equi-dispersion ($\alpha = 1$), or under-dispersion ($\alpha > 1$) of data.

A simple and efficient method for parameters estimation has been introduced in the present work and assessed using Monte Carlo simulation in a wide range of activity values.

The CMP distribution model herein proposed for PET data processing, has already been studied extensively in other disciplines such as Biology, Transportation, Marketing, eCommerce; to our knowledge, the use of CMP distribution model for characterizing PET data statistics remains unexplored. So the extension of the CMP model to analyze PET sinogram data is novel.

The proposed method has been applied both on simulated and experimental PET phantom data. The results of this study demonstrated that the CMP distribution could handle over-, equi-, and under-dispersed data with an adequate accuracy allowing to consider it a formidable tool in tracing the statistical behaviour of PET data during the several procedural corrections involved in PET reconstruction process.

It may be usefully implemented in image reconstruction algorithms and quantitative PET data analysis, especially in low counting emission data, as in dynamic PET data, where the method demonstrated the best accuracy.

References

-
- [1] M.Rahimi-Gorji, T.B.Gorji, M.Gorji-Bandpy. Details of regional particle deposition and airflow structures in a realistic model of human tracheobronchial airways: two-phase flow simulation, *Computers in Biology and Medicine*. 74:1 (2016) 1-17.
 - [2] M. Rahimi-Gorji, O. Pourmehran, M. Gorji-Bandpy, T.B. Gorji. CFD simulation of airflow behavior and particle transport and deposition in different breathing conditions through the realistic model of human airways, *Journal of Molecular Liquids*, 209 (2015) 121-133.
 - [3] F. Gibiino, F. Positano, V. Wiesinger, F. Giovannetti, G. Landini, L. Santarelli, MF(2013). Structured errors in reconstruction methods for Non-Cartesian MR data. *Computers in Biology and Medicine*, 43:12 (2013) 2256-2262.
 - [4] F. Gibiino, V. Positano, L. Landini, M.F. Santarelli. Regularization techniques on least squares non-uniform fast Fourier transform. *International Journal for Numerical Methods in Biomedical Engineering*. 29:5 (2013) 561 - 573
 - [5] Y.Vardi, L. A.Shepp, L.Kaufman. A Statistical Model for Positron Emission Tomography”, *Journal of the American Statistical Association*. 80:389 (1985) 8-20.
 - [6] B.Guérinand G.E. Fakhri. Realistic PET Monte Carlo Simulation With Pixelated Block Detectors, Light Sharing, Random Coincidences and Dead-Time Modeling. *IEEE Trans Nucl Sci*. 55:3 (2008) 942–952 .
 - [7] S.R.Cherry, J. A.Sorenson, M. E.Phelps. *Physics in nuclear medicine*. Elsevier Saunders Eds (2012).
 - [8] J.Qi and R. H.Huesman. Effect of Errors in the System Matrix on MAP Image Reconstruction. *Phys. Med Biol*. 50:14 (2005) 3297–3312.
 - [9] T.Teymurazyan, H.S.Riauka, D. R.Jans. Properties of Noise in Positron Emission Tomography Images Reconstructed with Filtered-Backprojection and Row-Action Maximum Likelihood Algorithm. *J Digit Imaging*. 26 (2013) 447–456.
 - [10] H. Zaidi (ed). *Quantitative Analysis in Nuclear Medicine imaging*. New York: Springer (2006).
 - [11] M.Dawood, X.Jiang and K.Schäfers (ed). *Correction Techniques in Emission Tomography*, Boca Raton, CRC Press, Series in Medical Physics and Biomedical Engineering (2012).

-
- [12] D.Brasse, P.E.Kinahan, C.Lartizien, C.Comtat, M.Casey and C.Michel. Correction Methods for Random Coincidences in Fully 3D Whole-Body PET: Impact on Data and Image Quality. *J Nucl Med* 46 (2005) 859–867.
- [13] M.L. Larsenand, A.B.Kostinski. Simple dead-time corrections for discrete time series of non-Poisson data. *Meas. Sci. Technol.* 20 (2009) 1-10.
- [14] F.Luisier, T.Blu, and M.Unser. Image Denoising in Mixed Poisson–Gaussian Noise. *IEEE Transactions On Image Processing*, 20:3 (2011) 696-708.
- [15] S.Ahn, and J. A.Fessler.Emission Image Reconstruction for Randoms-Precorrected PET Allowing Negative Sinogram Values, *IEEE Transactions on Medical Imaging*, 23:5 (2004) 591-601.
- [16] M. E.Casey. Point Spread Function Reconstruction in PET, Siemens Medical Solutions USA (2007).
- [17] Y.Jian, B.Planeta, R E.Carson. Evaluation of bias and variance in low-count OSEM list mode reconstruction. *Phys. Med. Biol.* 60 (2015) 15–29.
- [18] A. Reilhac, S. Tomeï, I. Buvat, C. Michel, F.Keheren, N.Costes. Simulation-based evaluation of OSEM iterative reconstruction methods in dynamic brain PET studies. *Neuroimage.* 39:1 (2008) 359-68.
- [19] M. D.Walker, M. C.Asselin, P. J.Julyan, M.Feldmann, P. S.Talbot, T.Jones, J. C.Matthews. Bias in iterative reconstruction of low-statistics PET data: benefits of a resolution model. *Phys. Med. Biol.* 56 (2011) 931–949.
- [20] G.L. Zeng. The ML-EM Algorithm is Not Optimal for Poisson Noise. *IEEE Transactions on Nuclear Science.* 62:5 (2015) 2096 - 2101.
- [21] J. D. Castillo and M. Pérez-Casany, Overdispersed and underdispersed Poisson generalizations, *Journal of statistical Planning and Inference*, 134 (2005), 486-500.
- [22] Krestyannikov E., Ruotsalainen U. (2004). Quantitatively accurate data recovery from attenuation-corrected sinogram using filtering of sinusoidal trajectory signals. *IEEE Nuclear Science Symposium*, 5, 3195-3199.
- [23] M.F. Santarelli, V. Positano, L. Landini. Measured PET Data Characterization with the Negative Binomial Distribution Model. *Journal of Medical and Biological Engineering.* (2016) In Press
- [24] R.A. Francis , S.R. Geedipally, S.D. Guikema, S.S. Dhavala, D. Lord , S. La Rocca . Characterizing the performance of the Conway-Maxwell Poisson generalized linear model. *Risk Anal.* 32:1 (2012) 167-83.
- [25] K.F. Sellers, G. Shmueli. Data Dispersion: Now You See It... Now You Don't. *Communications in Statistics - Theory and Methods.* 42:17 (2013) 3134-3147.
- [26] K.F. Sellers, S.Borle, G.Shmueli. The COM-Poisson Model for Count Data: A Survey of Methods and Applications. *Applied stochastic models in business and industry.* 28:2 (2012) 104-116.

-
- [27] M.Asalam, M.Azamb and C.H. Junc. A Control Chart for COM–Poisson Distribution Using Resampling and Exponentially Weighted Moving Average. *Qual. Reliab. Engng. Int.* 32 (2016) 727–735.
- [28] P.Sur, G.Shmueli, S.Bose & P.Dubey. Modeling Bimodal Discrete Data Using Conway-Maxwell-Poisson Mixture Models. *Journal of Business & Economic Statistics*, 33:3 (2015) 352-365.
- [29] M.S.Ridout, P.Besbeas. An empirical model for underdispersed count data. *Statistical Modelling*, 4 (2004) 77–89.
- [30] D.Lord, S.R.Geedipally, S.D.Gaikema. Extension of the application of Conway–Maxwell–Poisson models: analyzing traffic crash data exhibiting underdispersion. *Risk Analysis* . 30 (2010) 1268–1276.
- [31] P.Boatwright, S.Borle, J.B. Kadane. A model of the joint distribution of purchase quantity and timing. *Journal of the American Statistical Association*. 98 (2003) 564–72.
- [32] S. Singh, S. Borle, D. Jain. A generalized framework for estimating customer lifetime value when customer lifetimes are not observed. *Quantitative Marketing and Economics*. 7 (2009) 181–205.
- [33] G. Shmueli, T. P. Minka, J. B. Kadane, S. Borle, and P. Boatwright. A Useful Distribution for Fitting Discrete Data: Revival of the Conway-Maxwell-Poisson Distribution. *Applied Statistics*, 54 (2005) 127-142.
- [34] A. Papoulis, *Probability, Random Variables, and Stochastic Processes* 4th edn. McGraw-Hill (2002).
- [35] D. R. Cox, P. A. W. Lewis. *The Statistical Analysis of Series of Events*. London: Methuen(1966).
- [36] Fano U. (1947). Ionization Yield of Radiations. II. The Fluctuations of the Number of Ions. *Phys. Rev.* 72:26.
- [37] R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria (2013) URL <http://www.R-project.org/>.
- [38] E. Krestyannikov, U. Ruotsalainen. Quantitatively accurate data recovery from attenuation-corrected sinogram using filtering of sinusoidal trajectory signals. *IEEE Nuclear Science Symposium*, 5 (2004) 3195-3199.
- [39] F. K. Sellers, S. Borle, and G. Shmueli. The COM-Poisson model for count data: a survey of methods and applications. *Appl. Stochastic Models Bus. Ind.*, 28 (2012) 104–116.

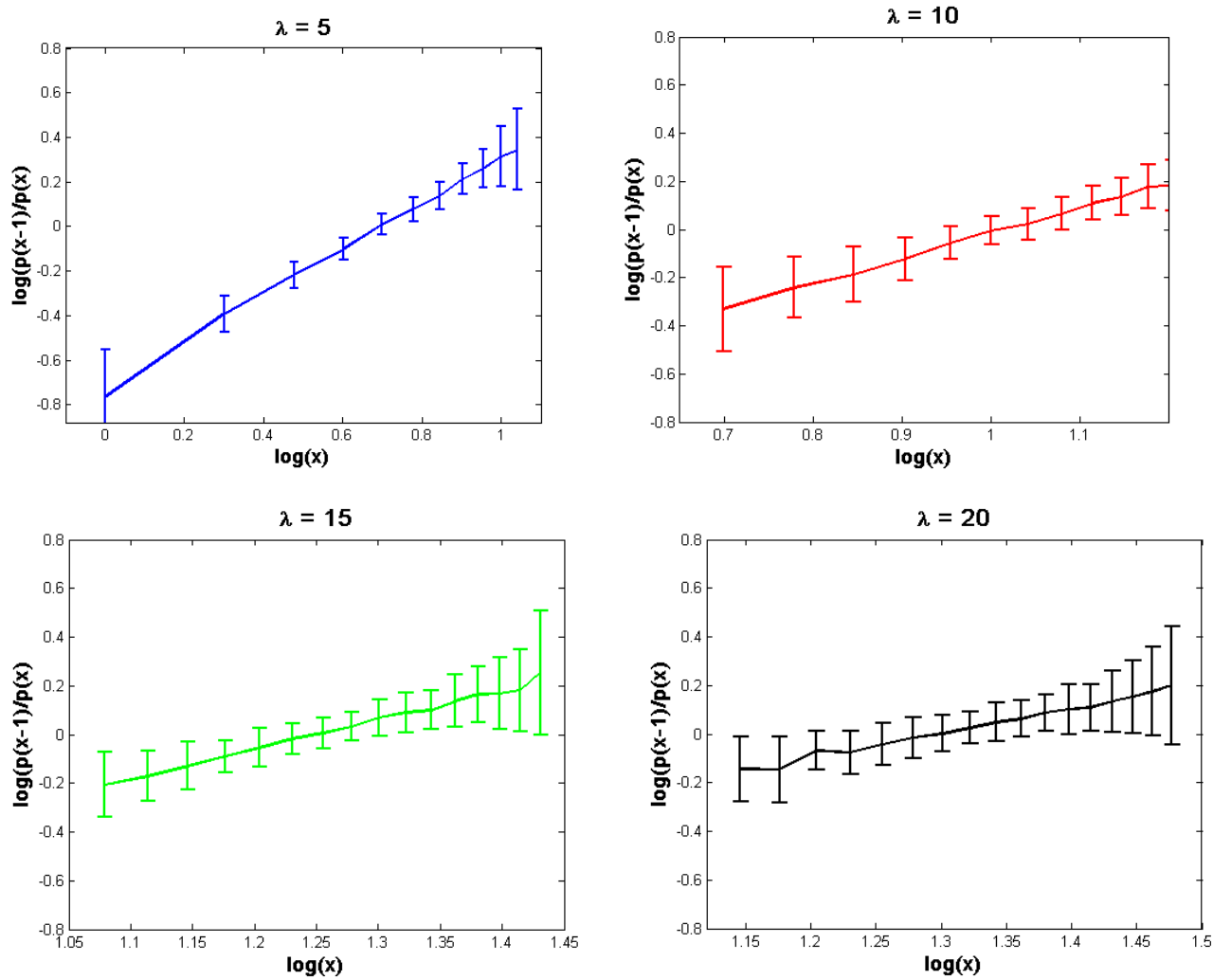


Figure 1. Typical trends of $\log(p(x-1)/p(x))$ from eq. (5) vs $\log(x)$ for four different centring parameter λ ($\lambda = 5, 10, 15, 20$) and dispersion parameter $\mathcal{V} = 1$.

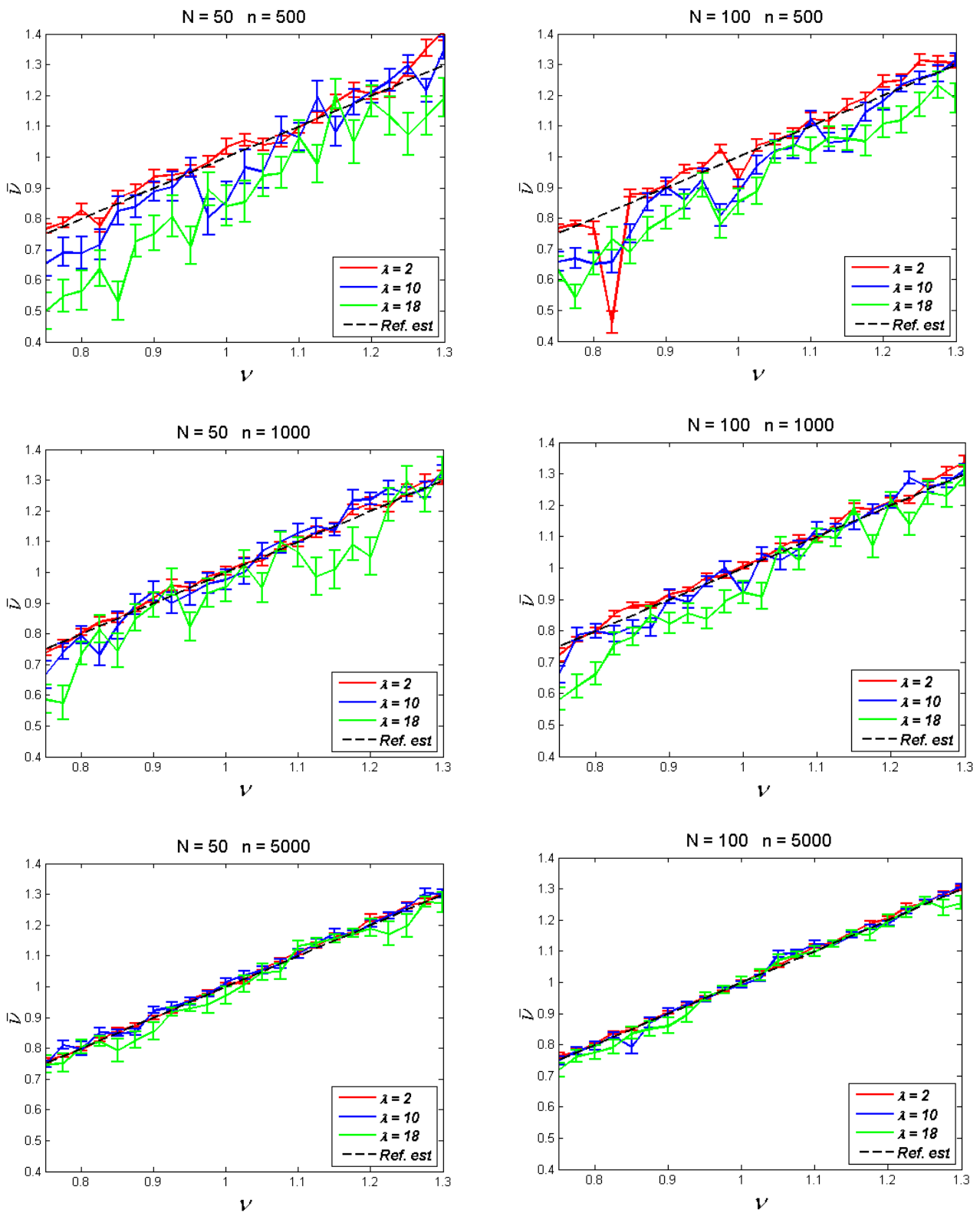


Figure 2. estimated dispersion parameter \bar{V} vs true value V (in log scale), for three different values of the centring parameter λ ($\lambda = 2, 10, 18$); sample size $n=500$ in the first line, $n=1000$ in the middle line, $n=5000$ in the last line; $N = 50$ in the left column, $N = 100$ in the right column. The dotted reference line is also shown, representing the true $\bar{V} = V$ regression line

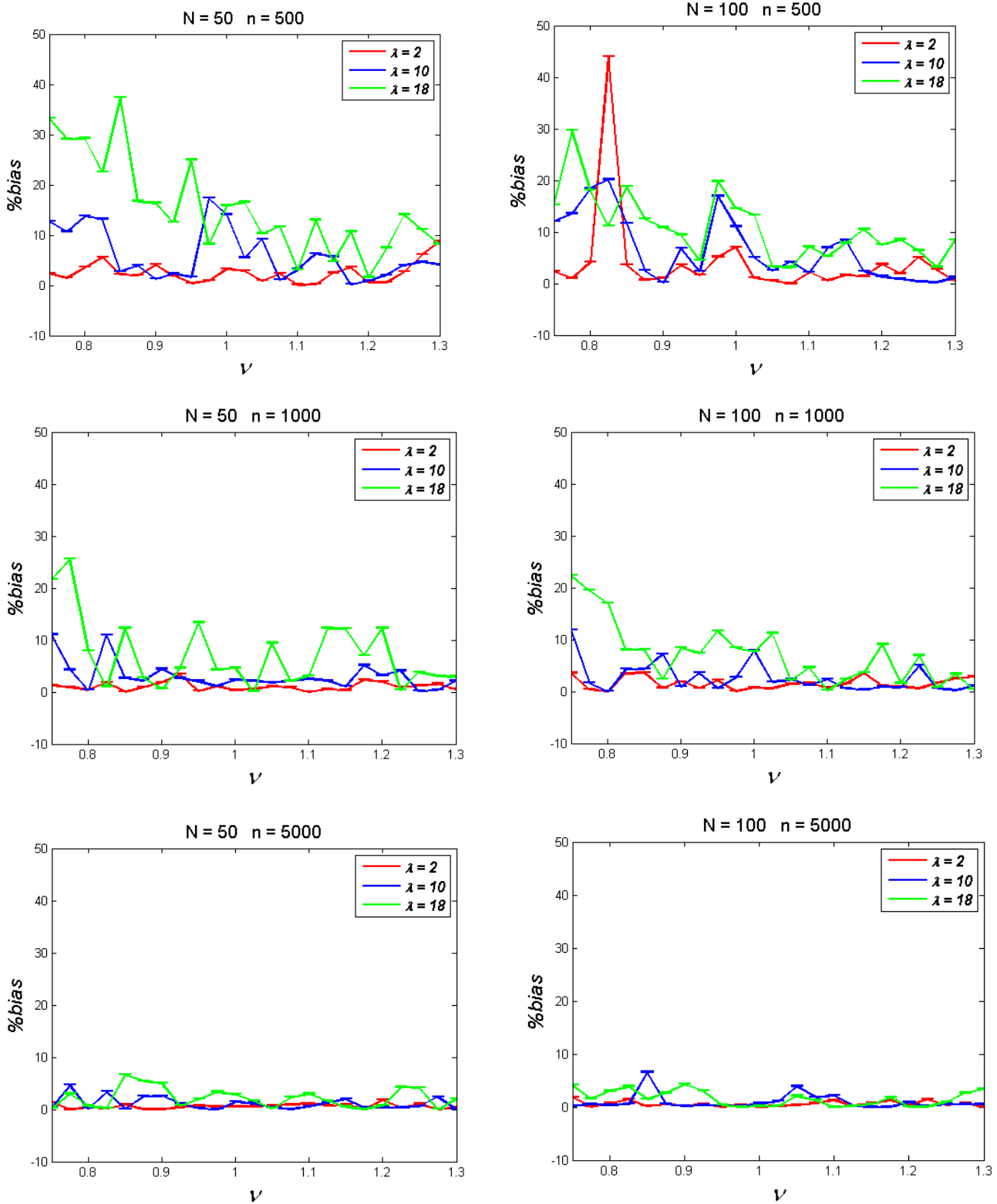


Figure 3. %bias on estimating the dispersion parameter values $\bar{\nu}$ vs true values ν (in log scale), for three different values of the centring parameter λ ($\lambda = 2, 10, 18$); sample size $n = 500$ in the first line, $n = 1000$ in the middle line, $n = 5000$ in the last line; $N = 50$ in the left column, $N = 100$ in the right column

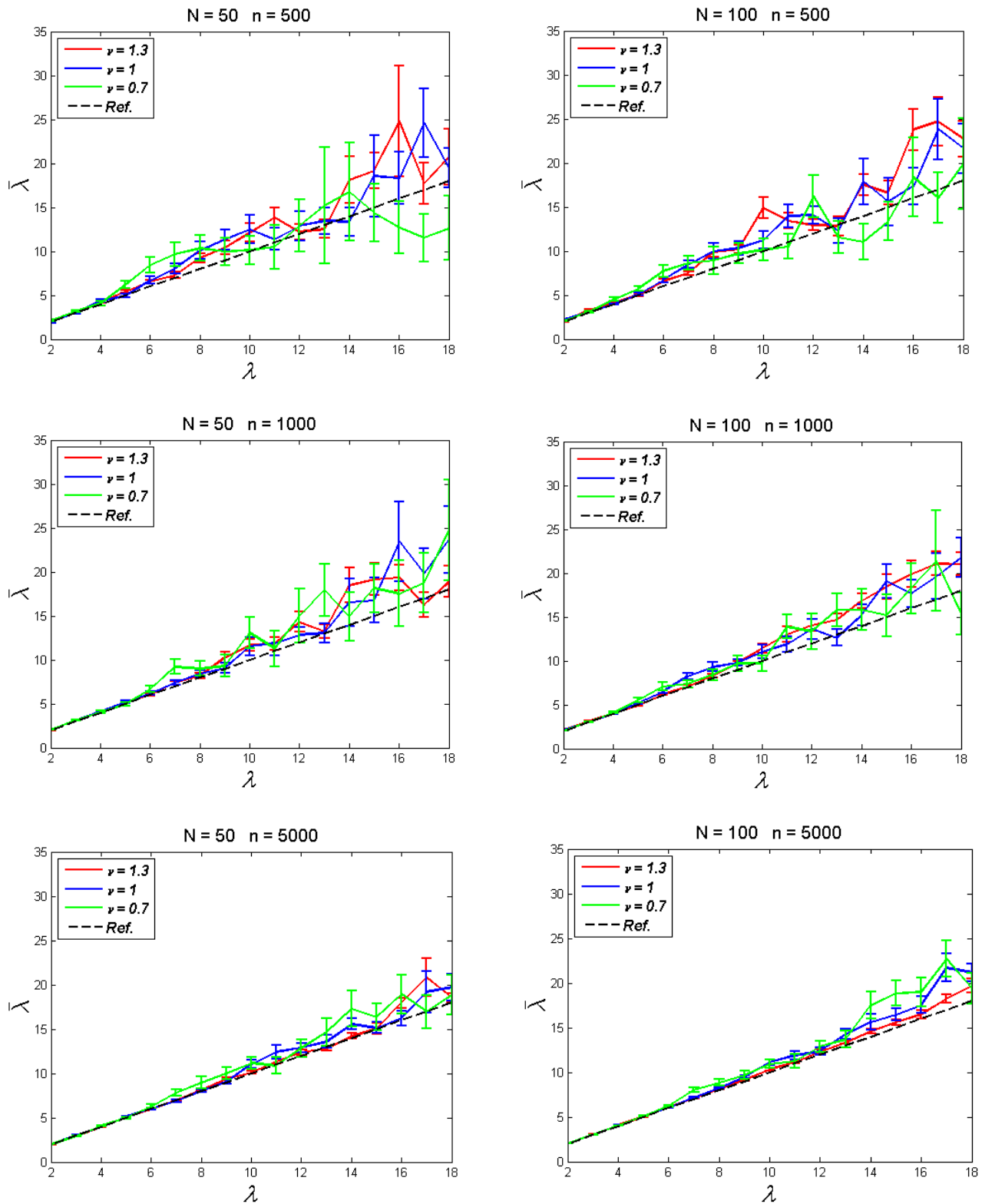


Figure 4. estimated centring parameter $\bar{\lambda}$ vs true values λ at three values of the dispersion parameter V ($V = 1.3, 1, 0.7$); sample size $n = 500$ in the first line, $n = 1000$ in the middle line, $n = 5000$ in the last line; $N = 50$ in the left column, $N = 100$ in the right column. The dotted reference line is also shown, representing the true $\bar{\lambda} = \lambda$ regression line

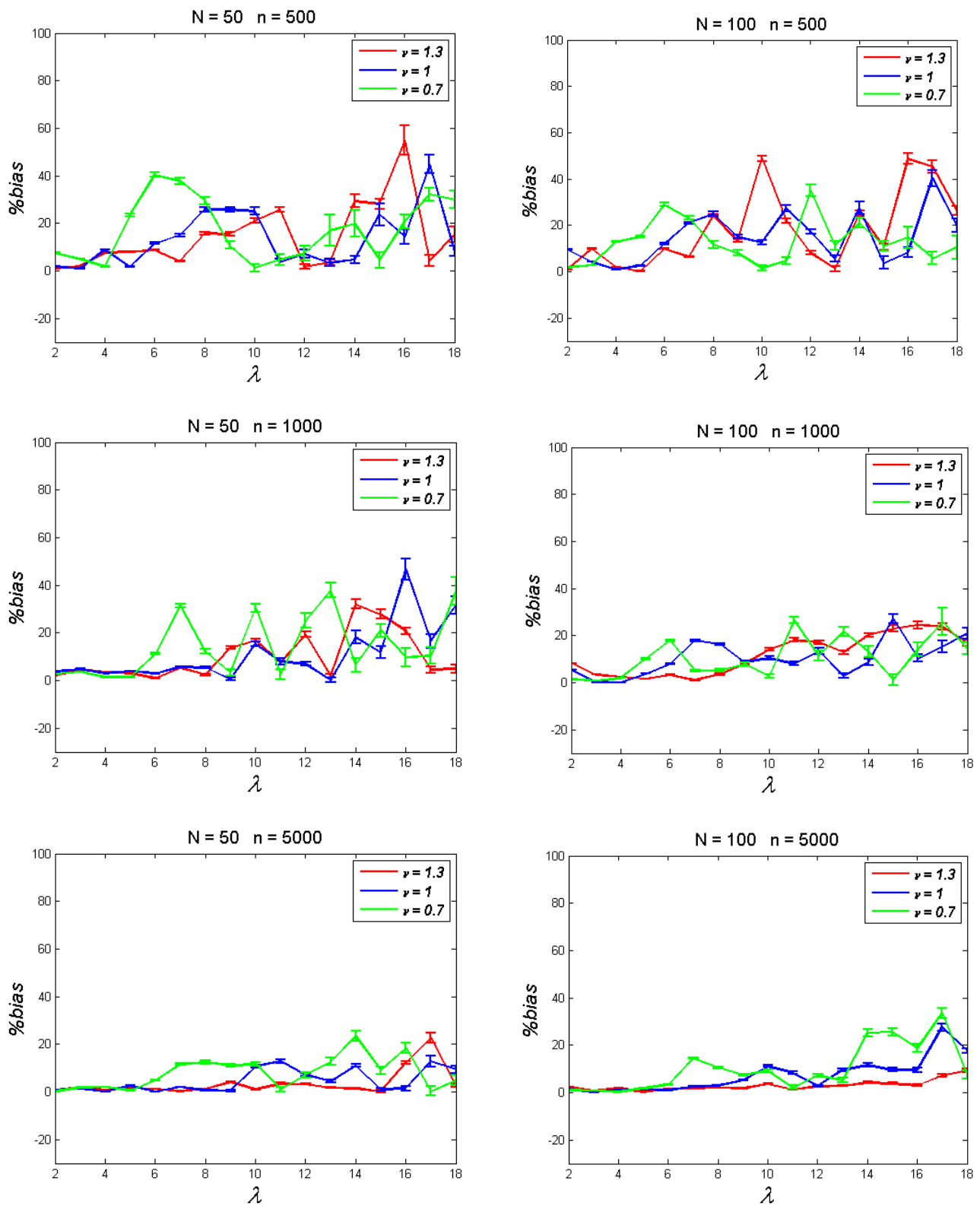


Figure 5. %bias on estimating the centring parameter values $\bar{\lambda}$ estimates vs true values λ at three values of the dispersion parameter ν ($\nu = 1.3, 1.0, 0.7$); sample size $n = 500$ in the first line, $n = 1000$ in the middle line, $n = 5000$ in the last line; $N = 50$ in the left column, $N = 100$ in the right column

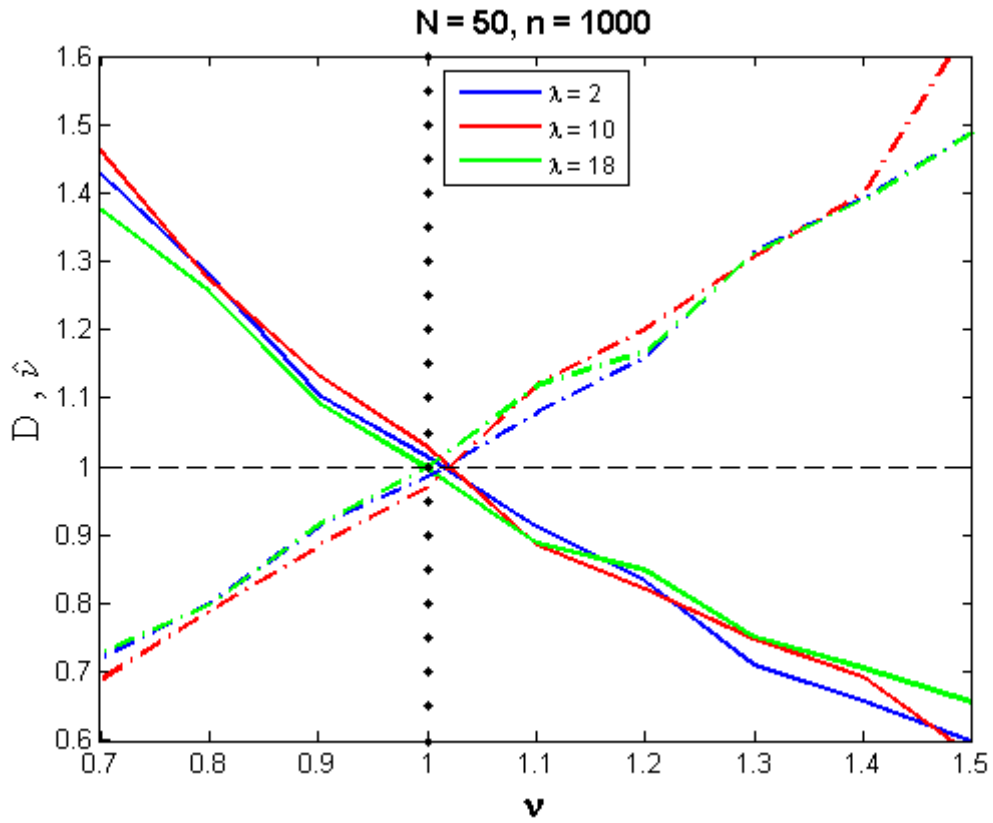
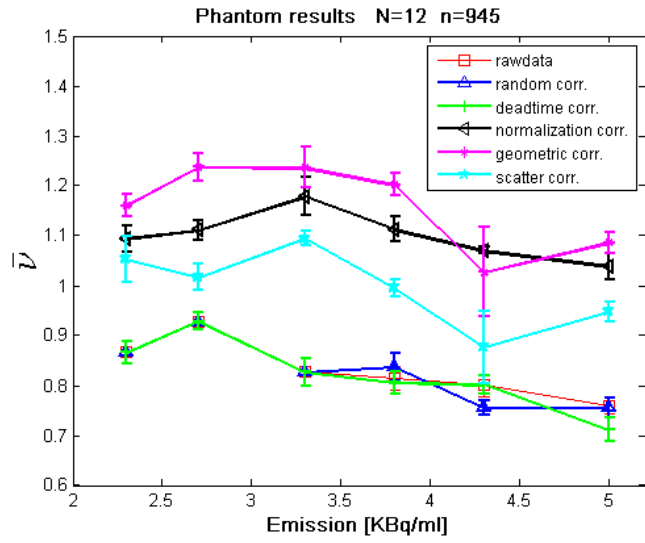
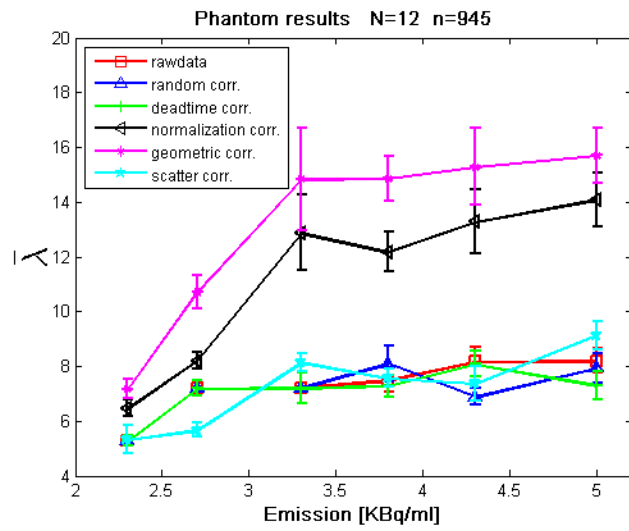


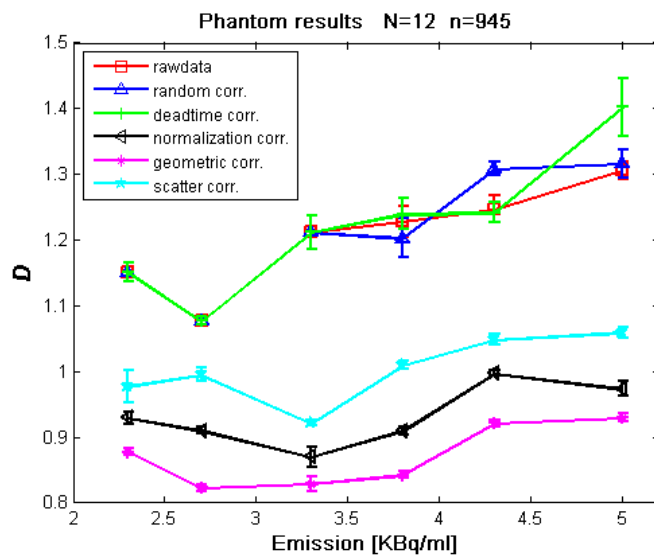
Figure 6. dispersion index D (continuous lines) and dispersion parameter \hat{v} (dotted lines) at three values of the centring parameter λ ($\lambda = 2, 10, 18$) as function of the true values V .



a



b



c

Figure 7. experimental results; a: dispersion parameter $\hat{\nu}$ estimates vs activity; b: centring parameter $\hat{\lambda}$ estimates vs activity; c: dispersion index D vs activity.

Figure 1a
[Click here to download high resolution image](#)

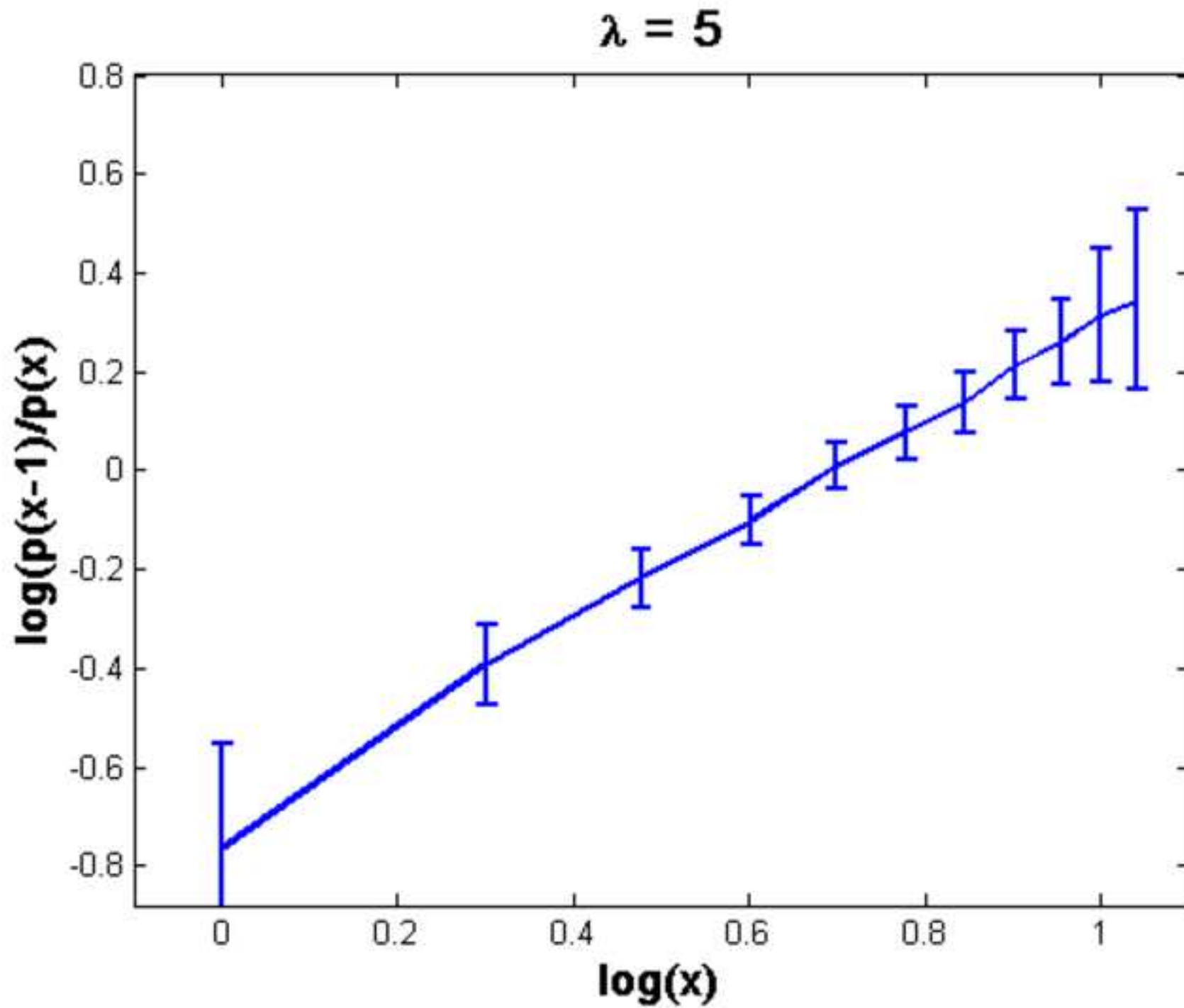


Figure 1b
[Click here to download high resolution image](#)

$\lambda = 10$

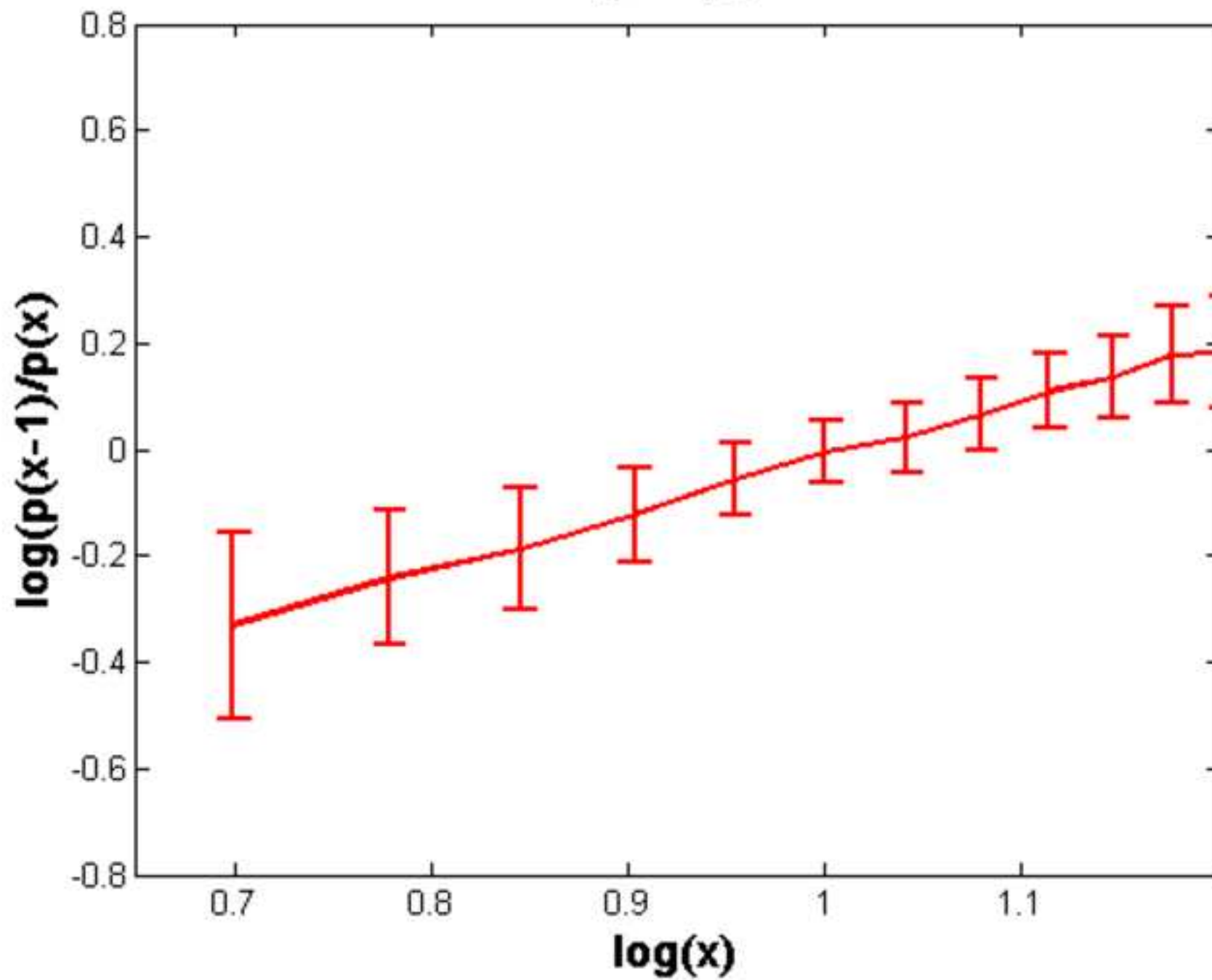


Figure 1c
[Click here to download high resolution image](#)

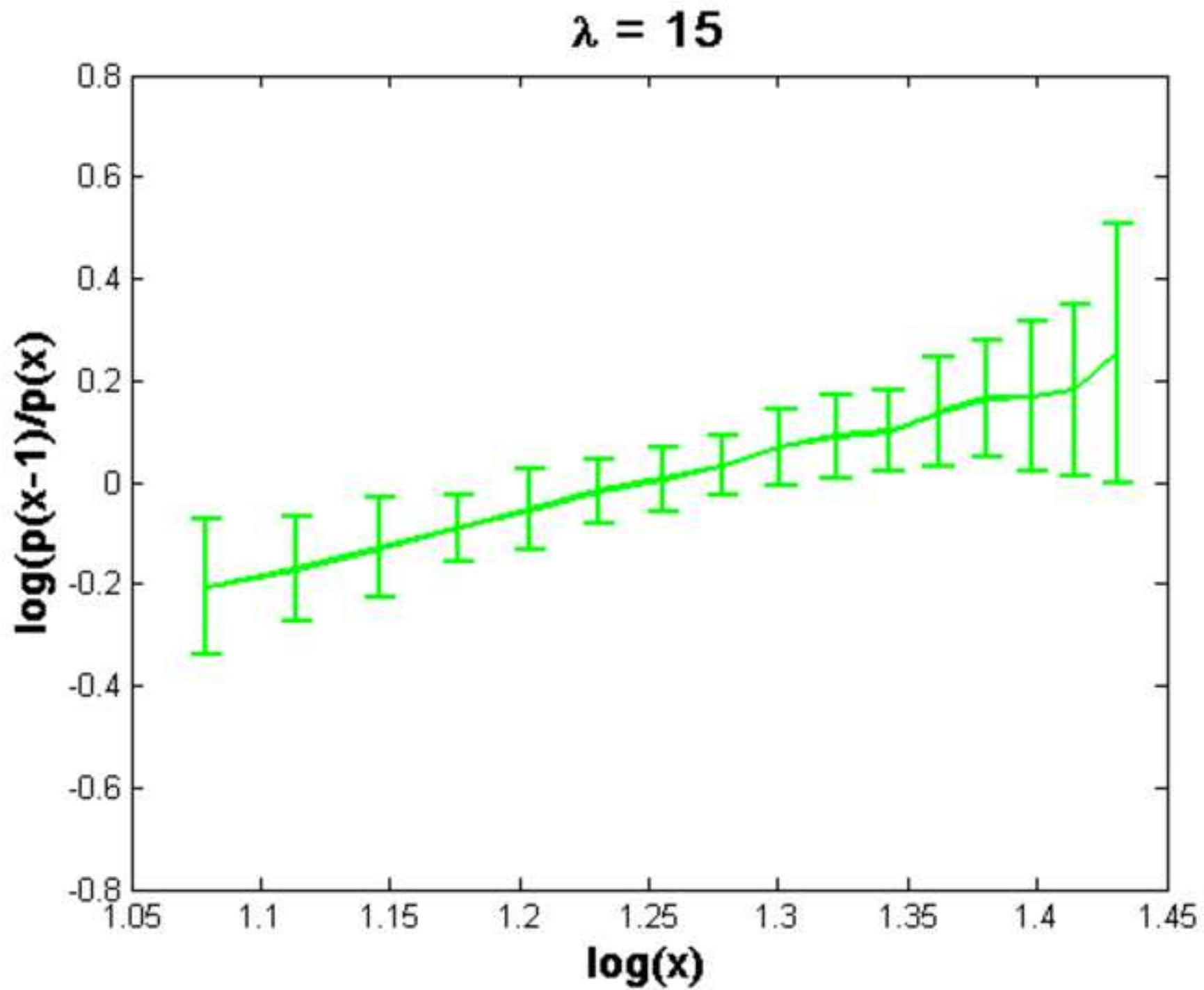


Figure 1d
[Click here to download high resolution image](#)

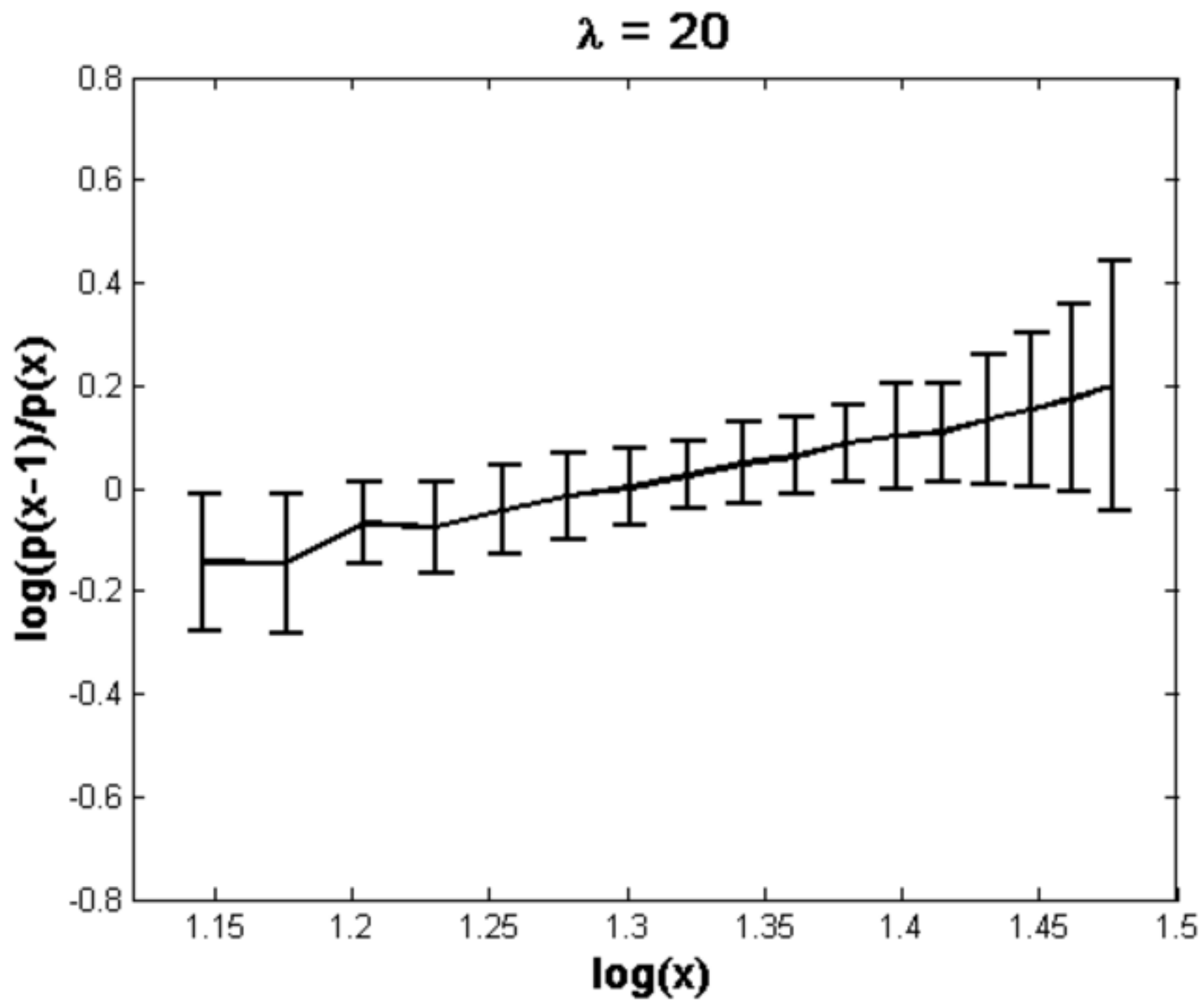


Figure 2a

[Click here to download high resolution image](#)

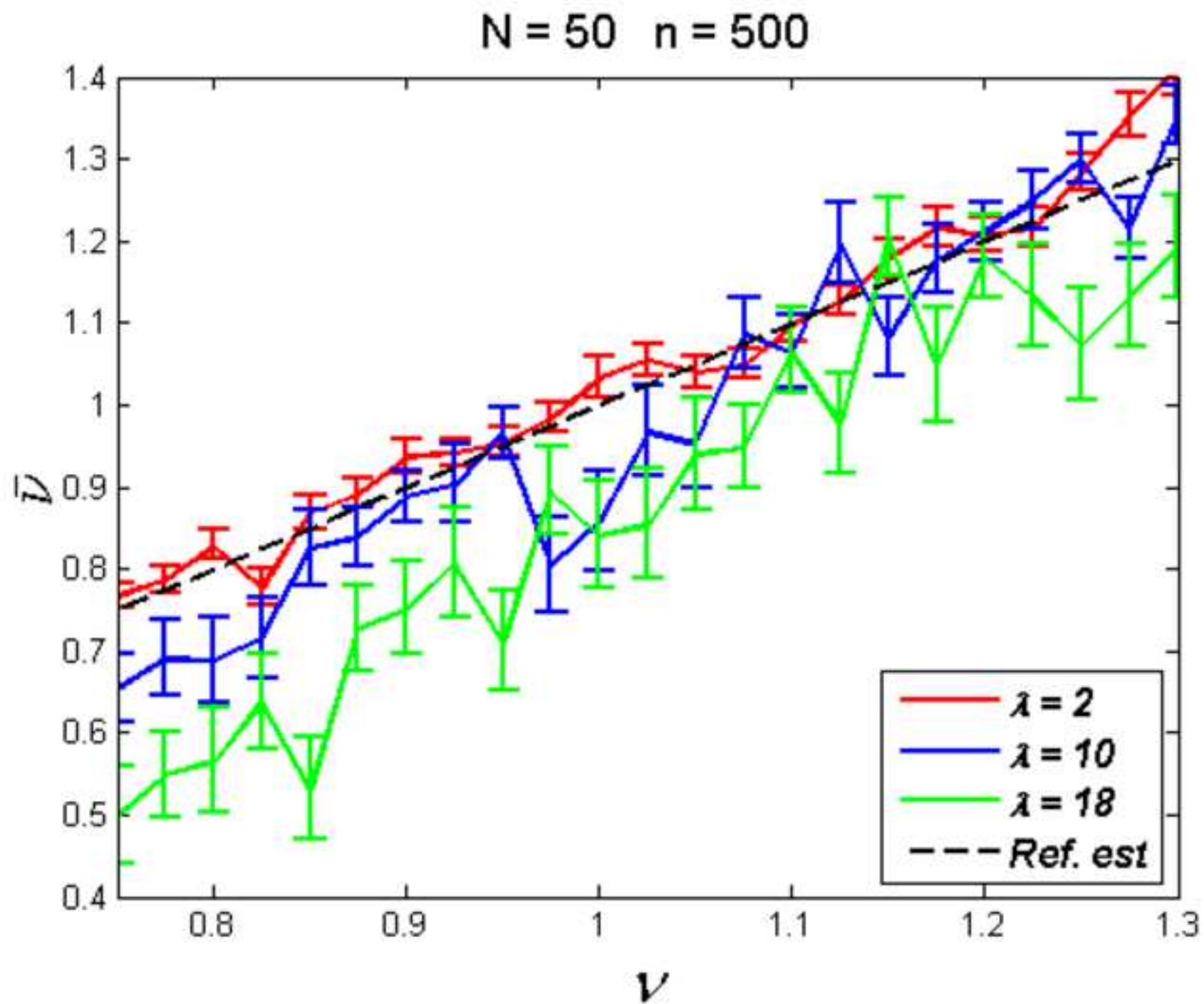


Figure 2b
[Click here to download high resolution image](#)

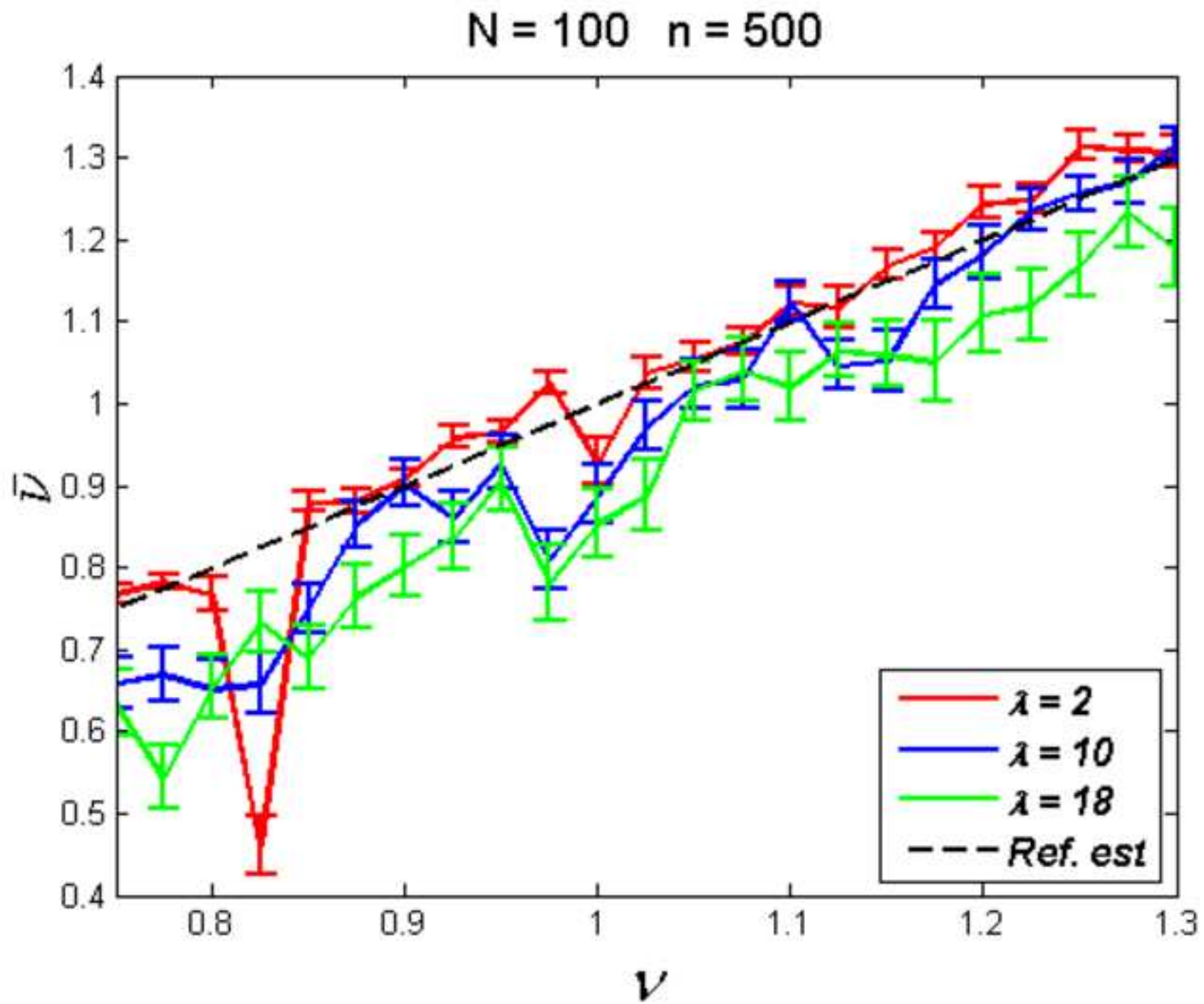


Figure 2c
[Click here to download high resolution image](#)

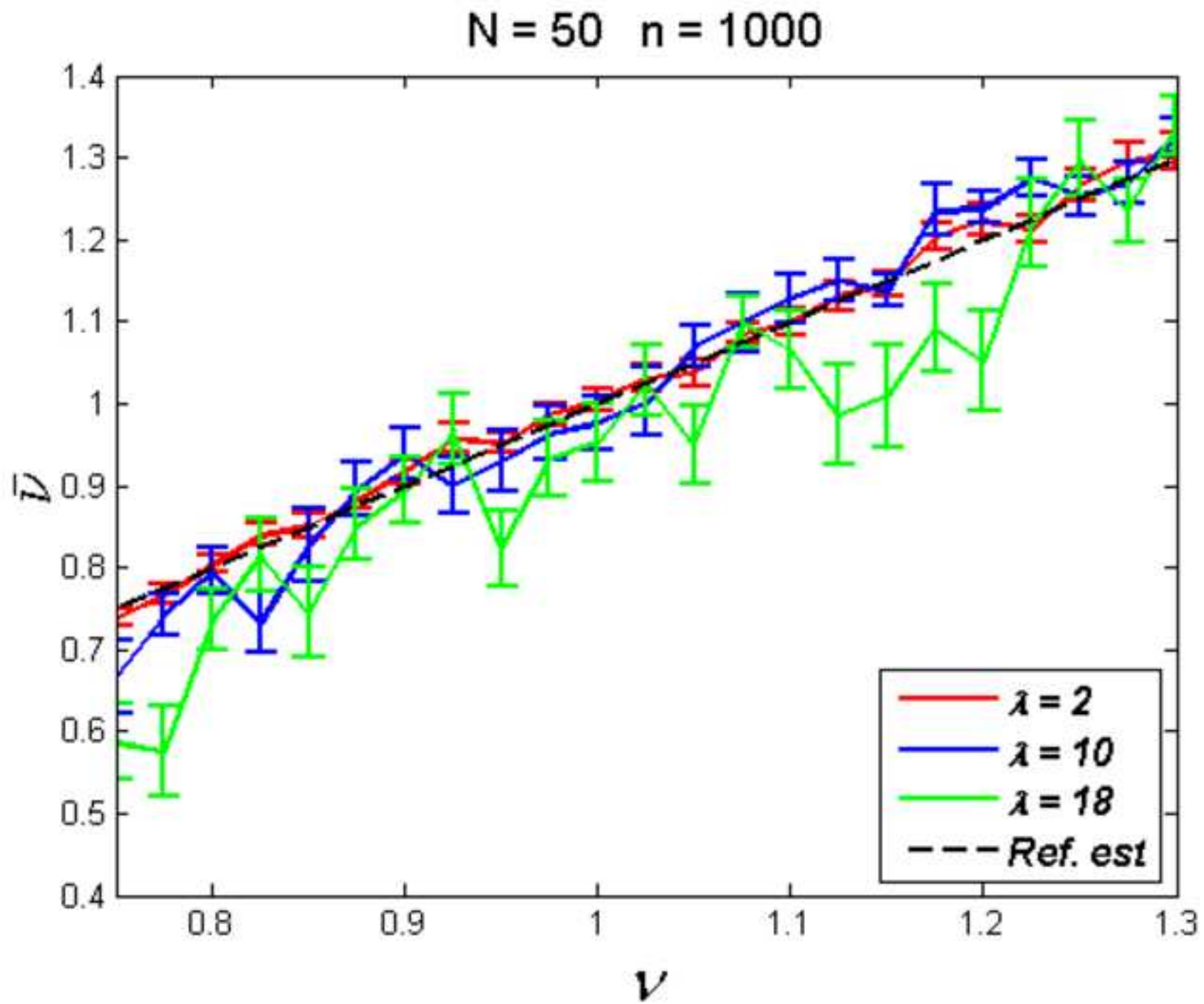


Figure 2d
[Click here to download high resolution image](#)

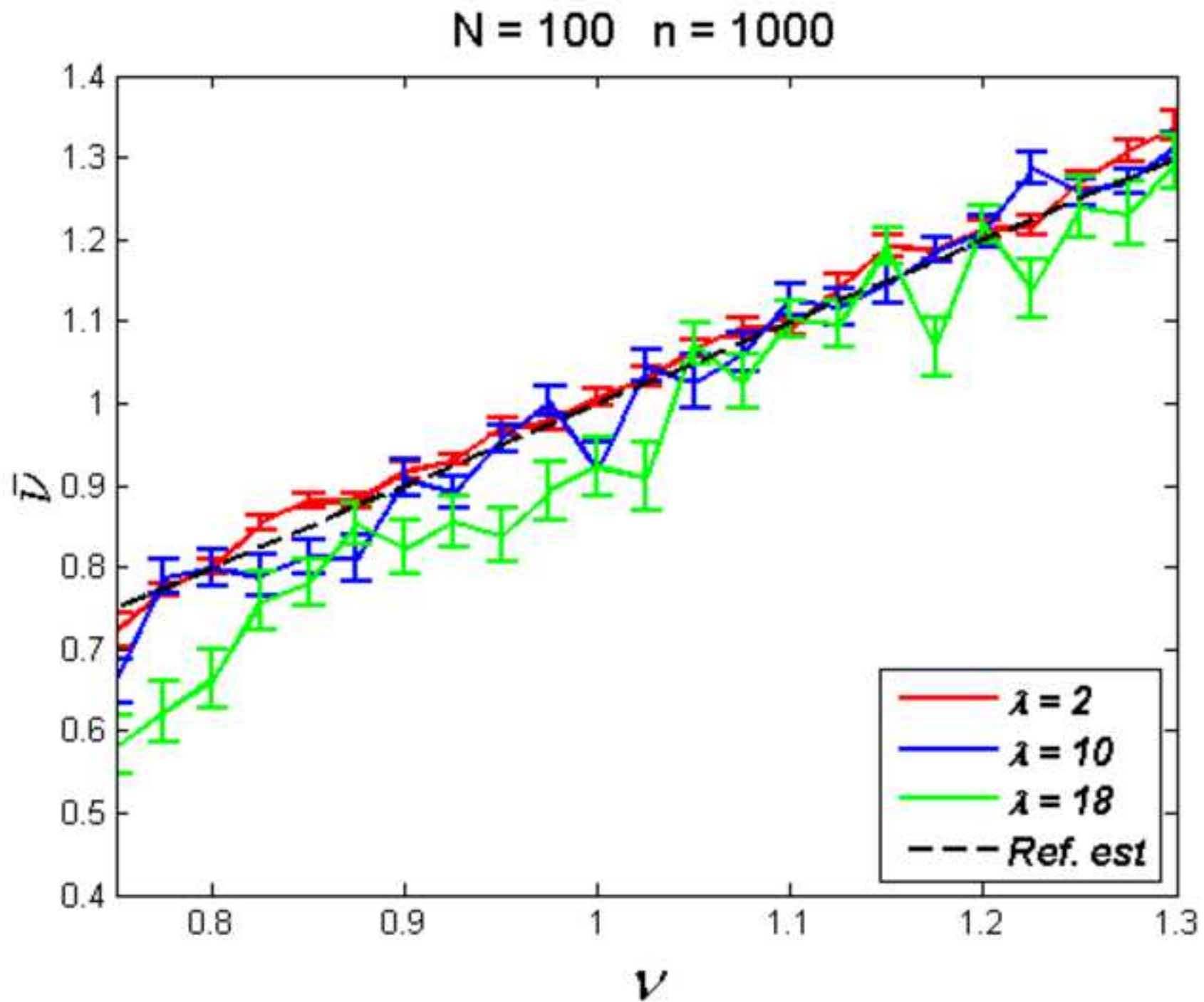


Figure 2e
[Click here to download high resolution image](#)

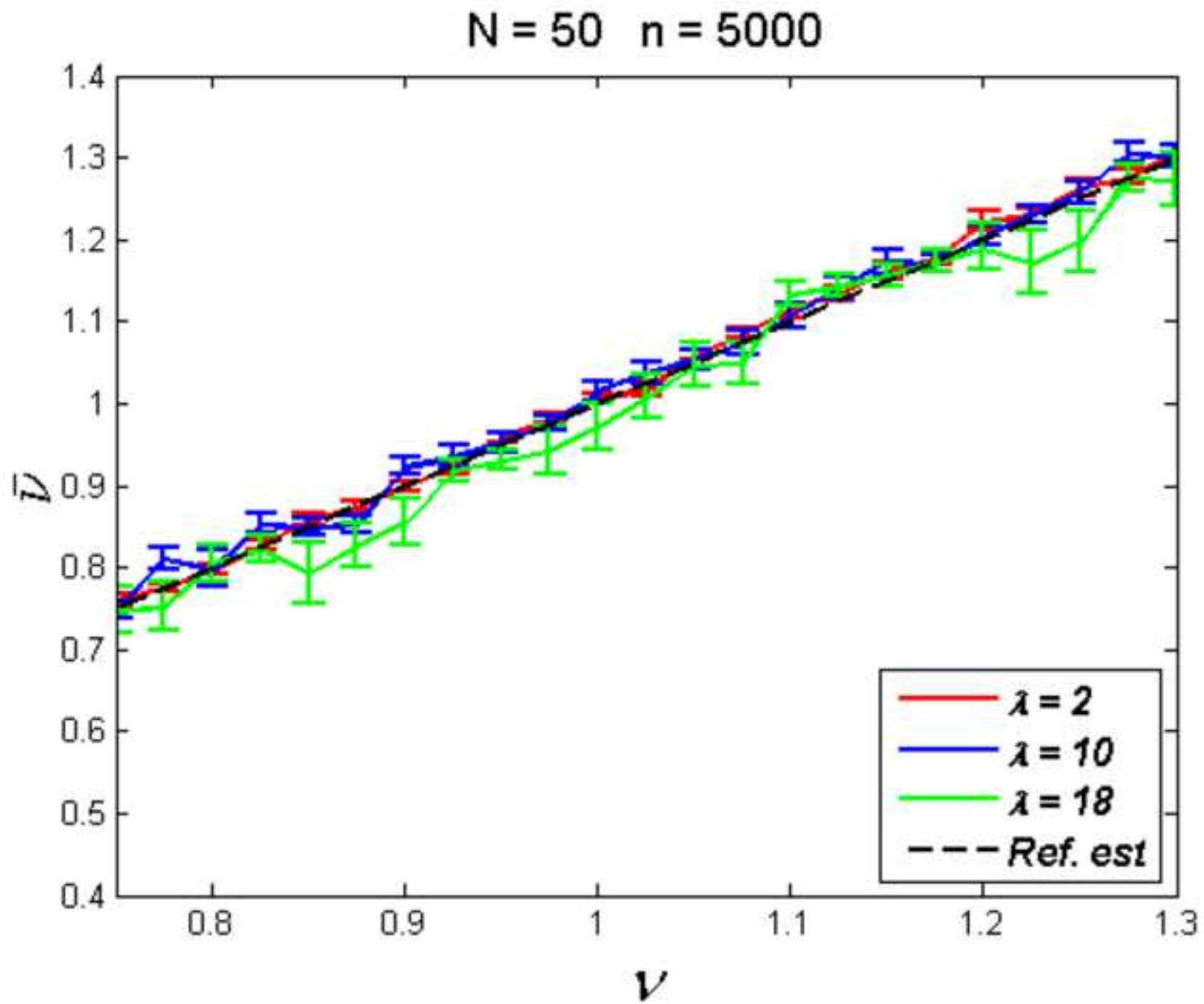


Figure 2f
[Click here to download high resolution image](#)

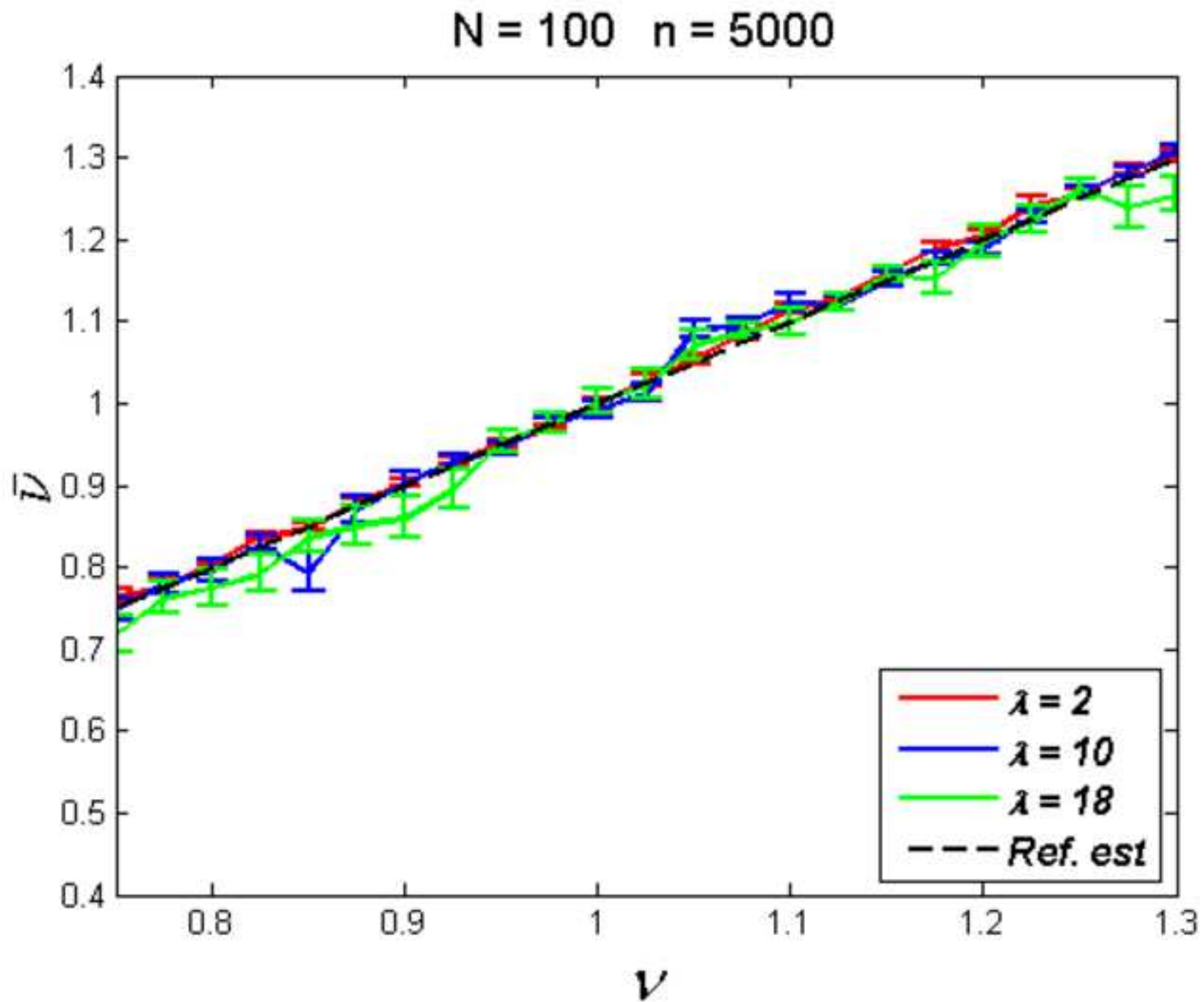


Figure 3a
[Click here to download high resolution image](#)

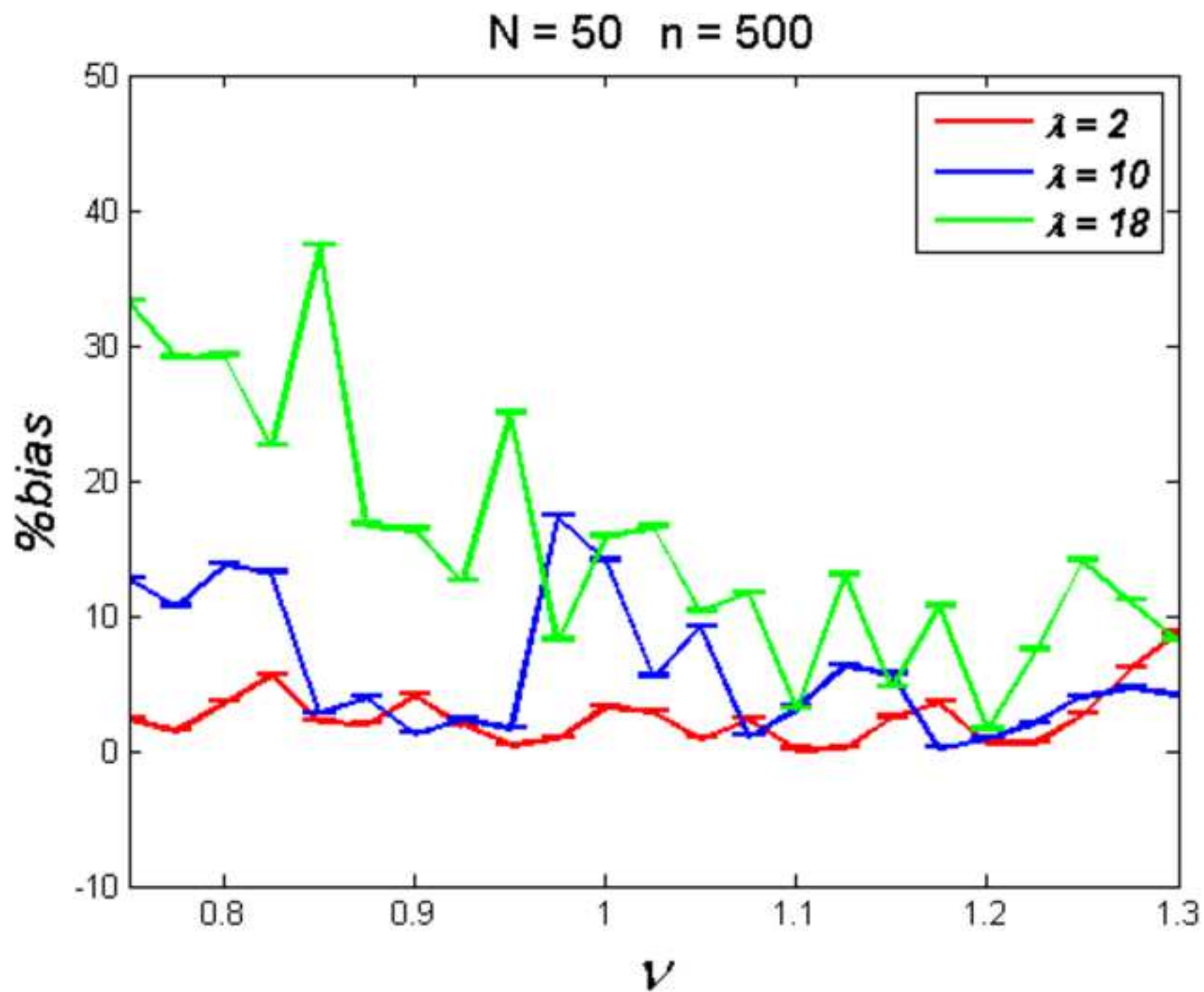


Figure 3b
[Click here to download high resolution image](#)

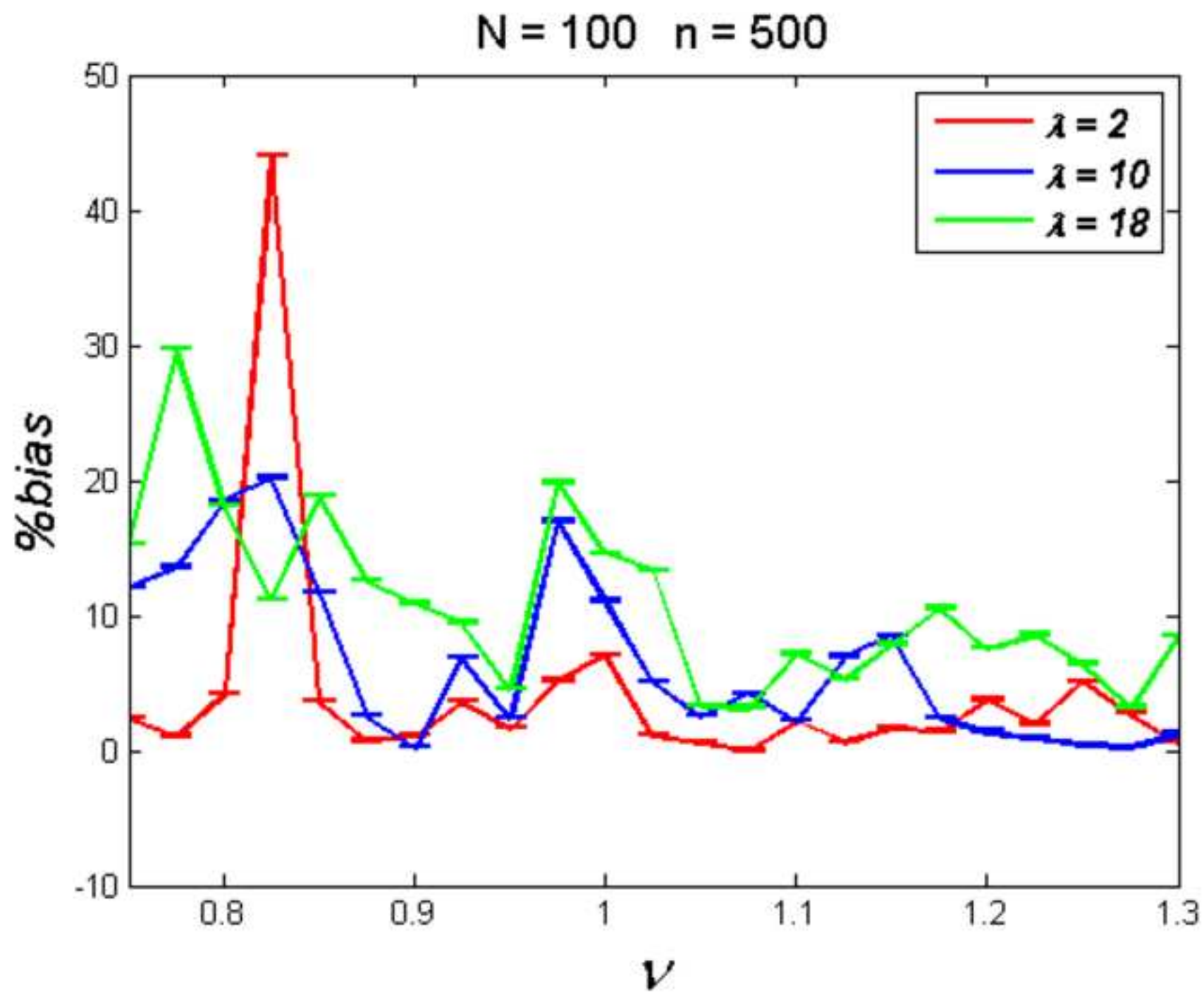


Figure 3c
[Click here to download high resolution image](#)

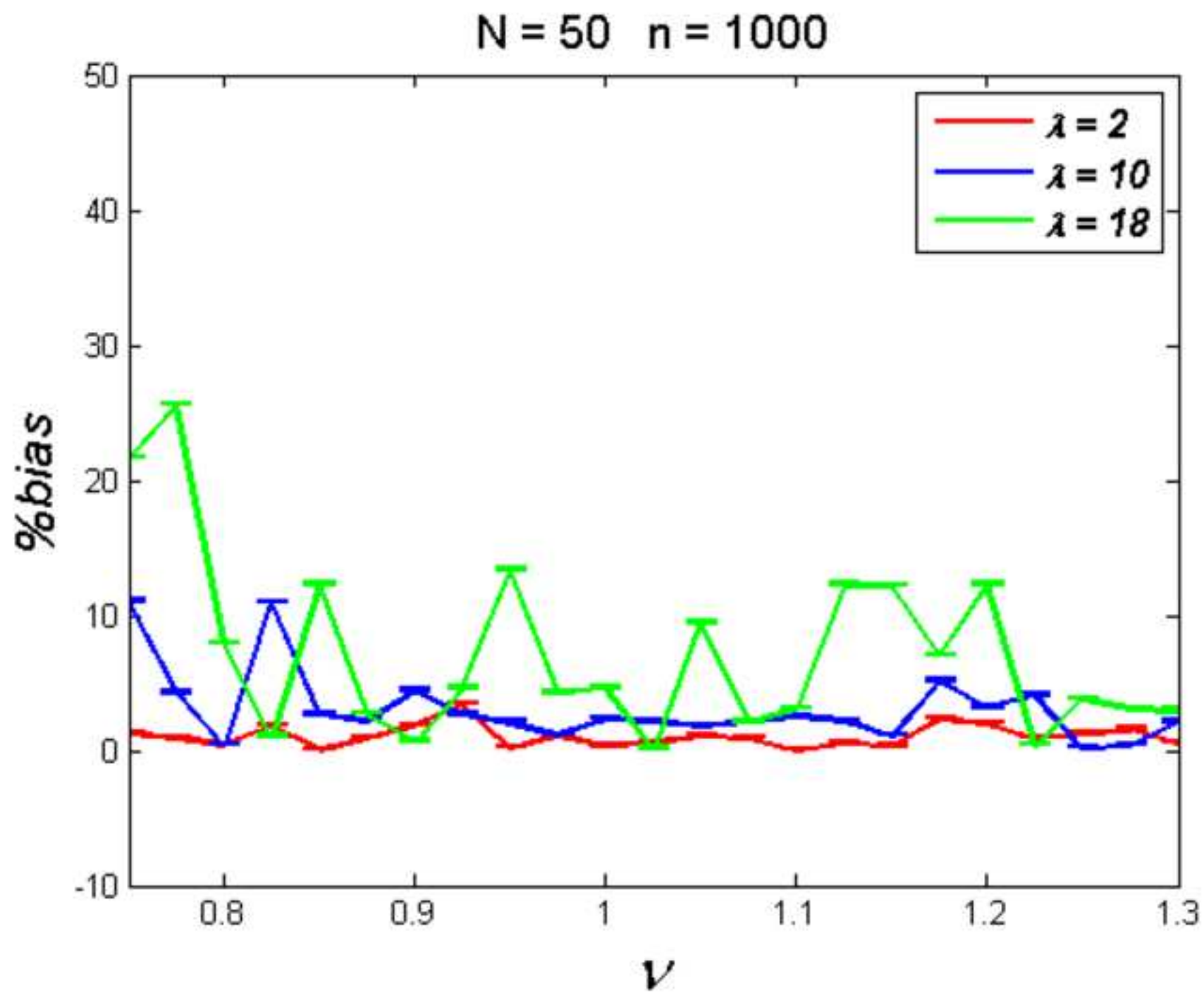


Figure 3d
[Click here to download high resolution image](#)

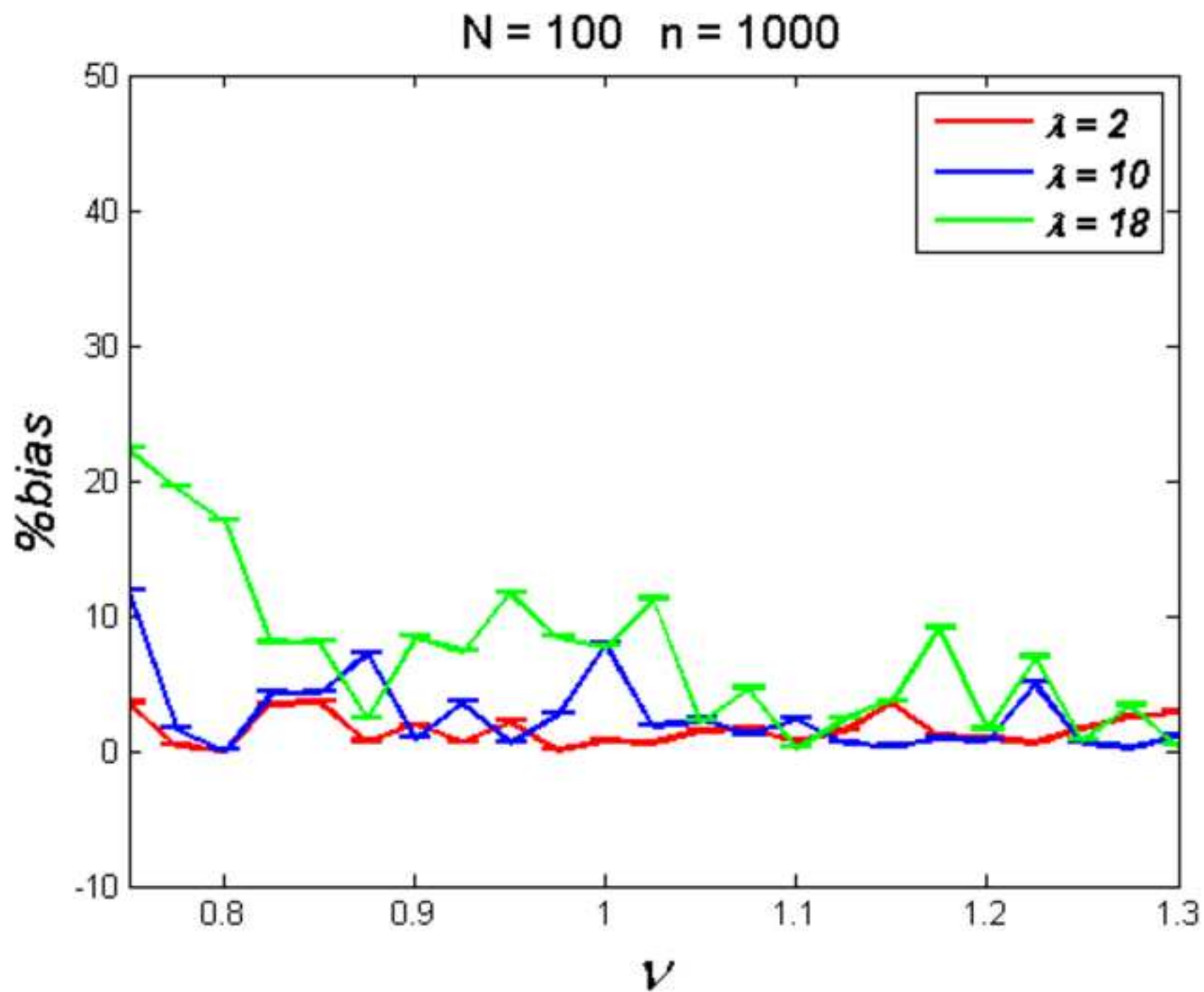


Figure 3e
[Click here to download high resolution image](#)

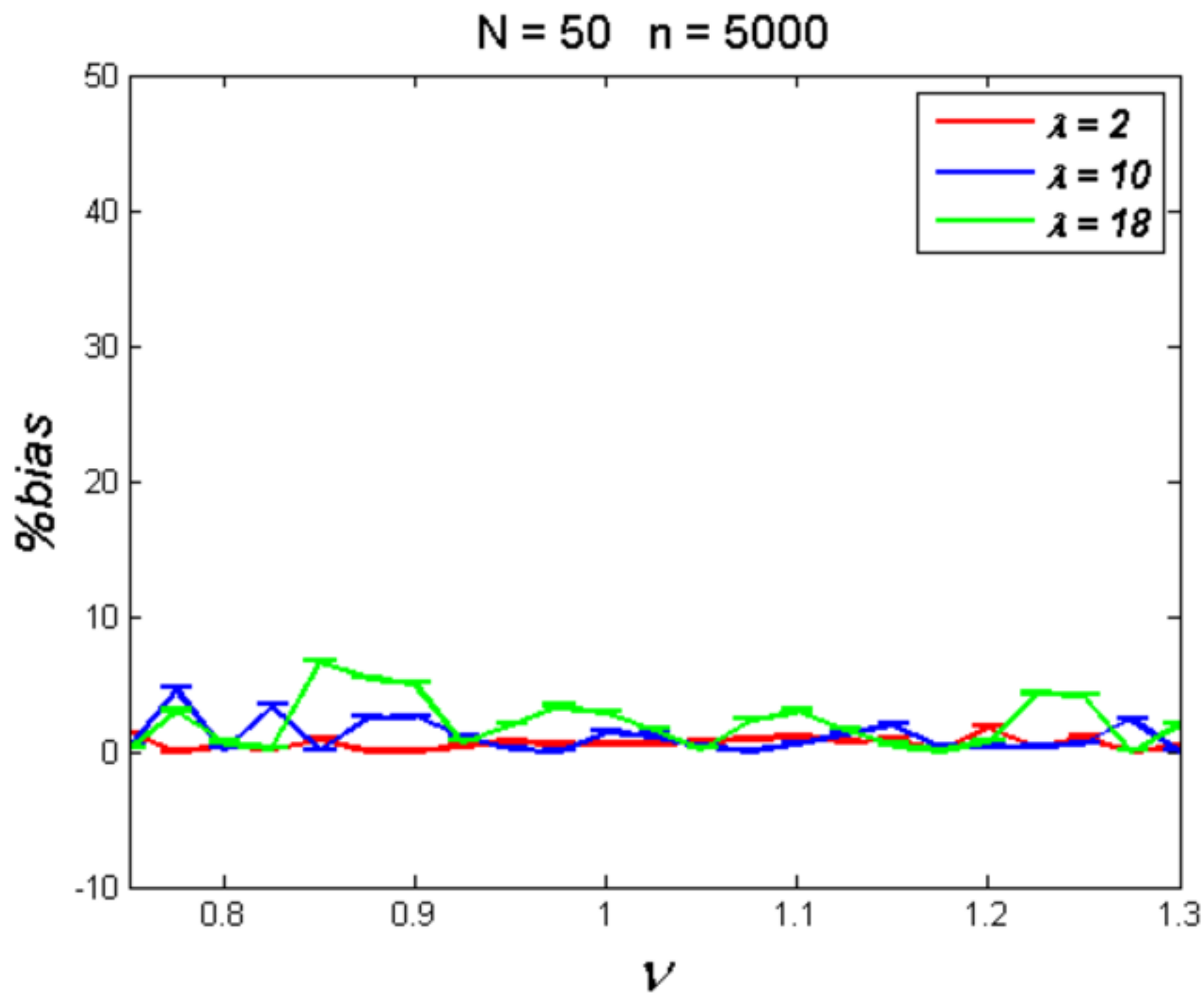


Figure 3f
[Click here to download high resolution image](#)

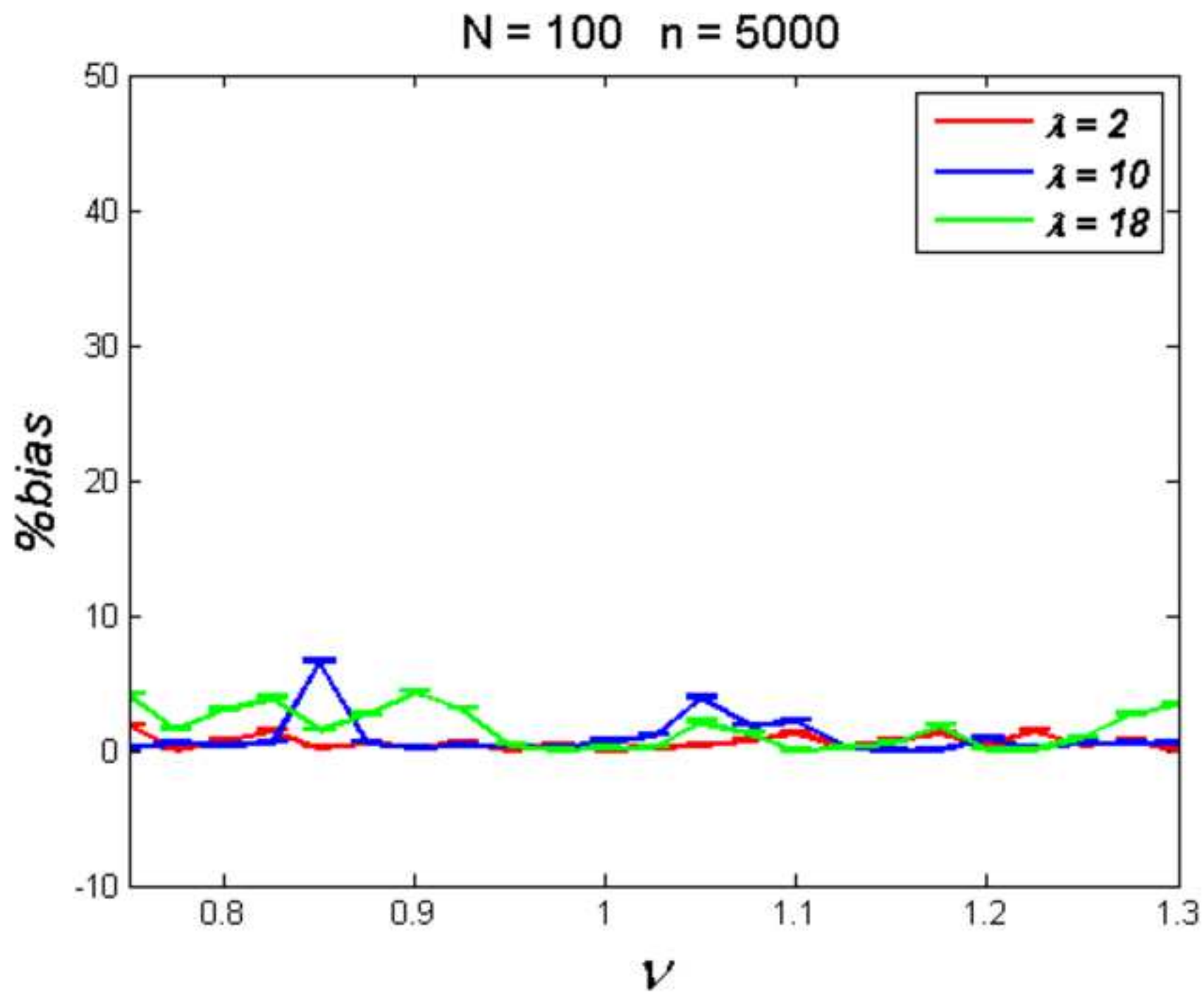


Figure 4a
[Click here to download high resolution image](#)

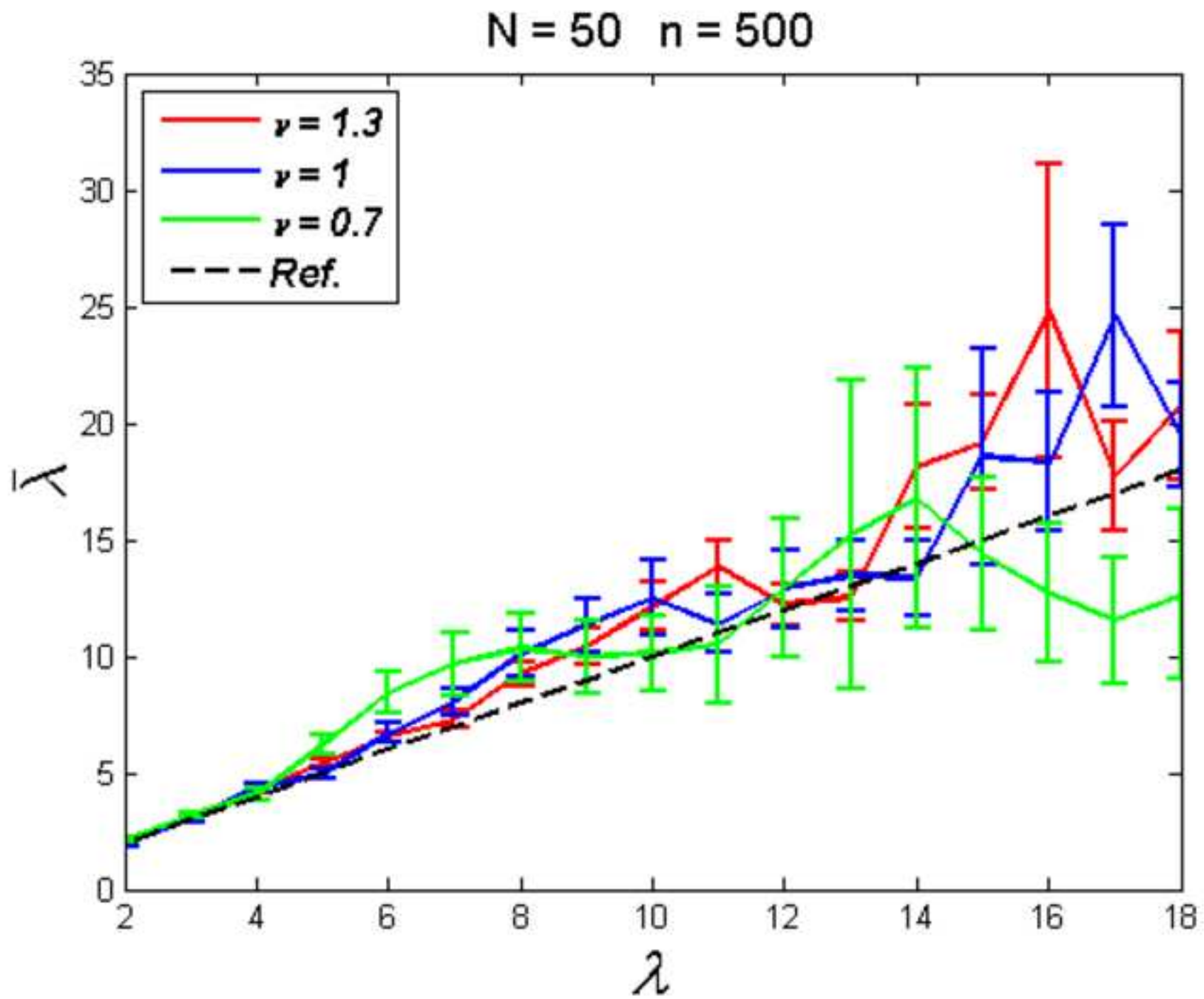


Figure 4b
[Click here to download high resolution image](#)

$N = 100$ $n = 500$

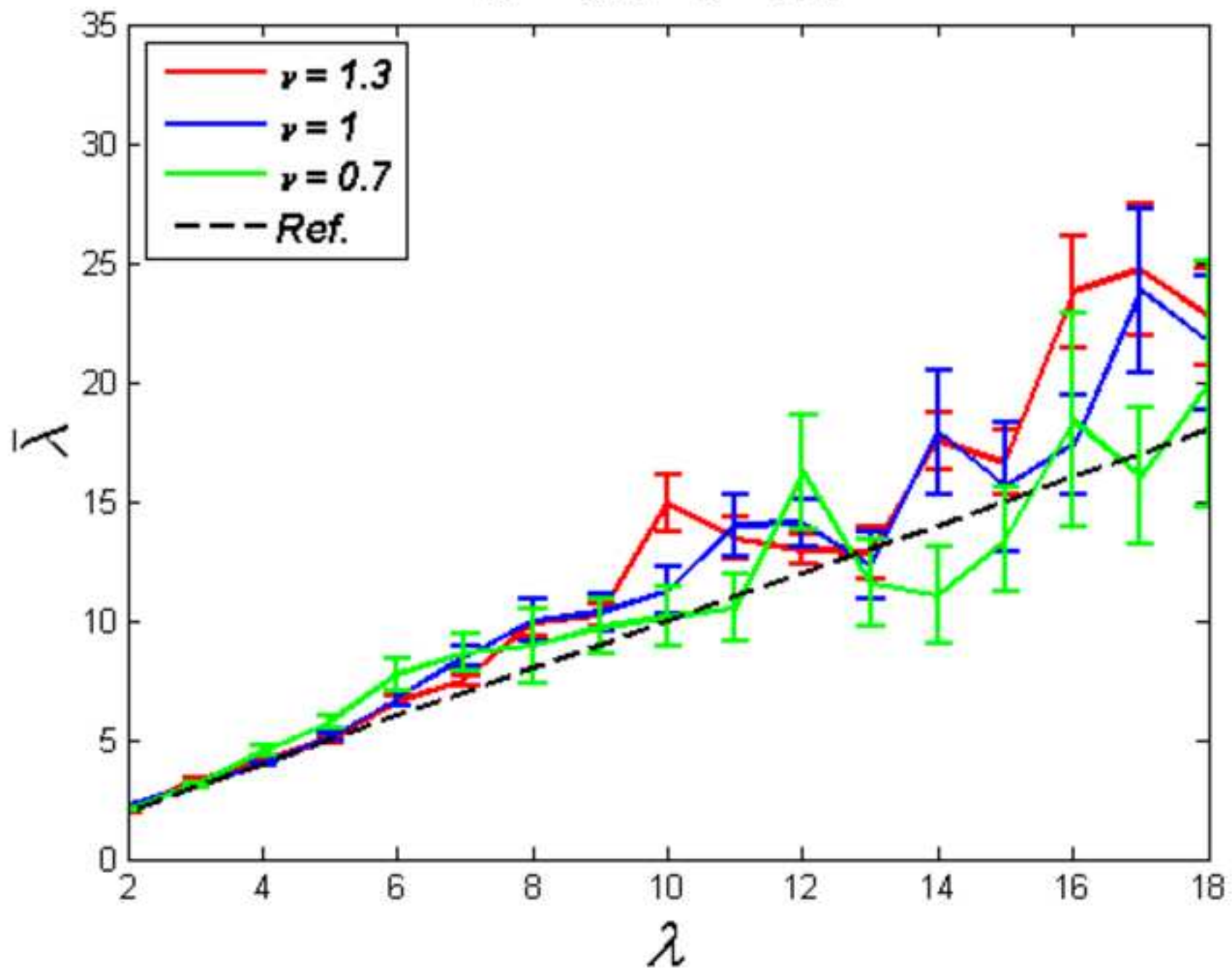


Figure 4c
[Click here to download high resolution image](#)

$N = 50$ $n = 1000$

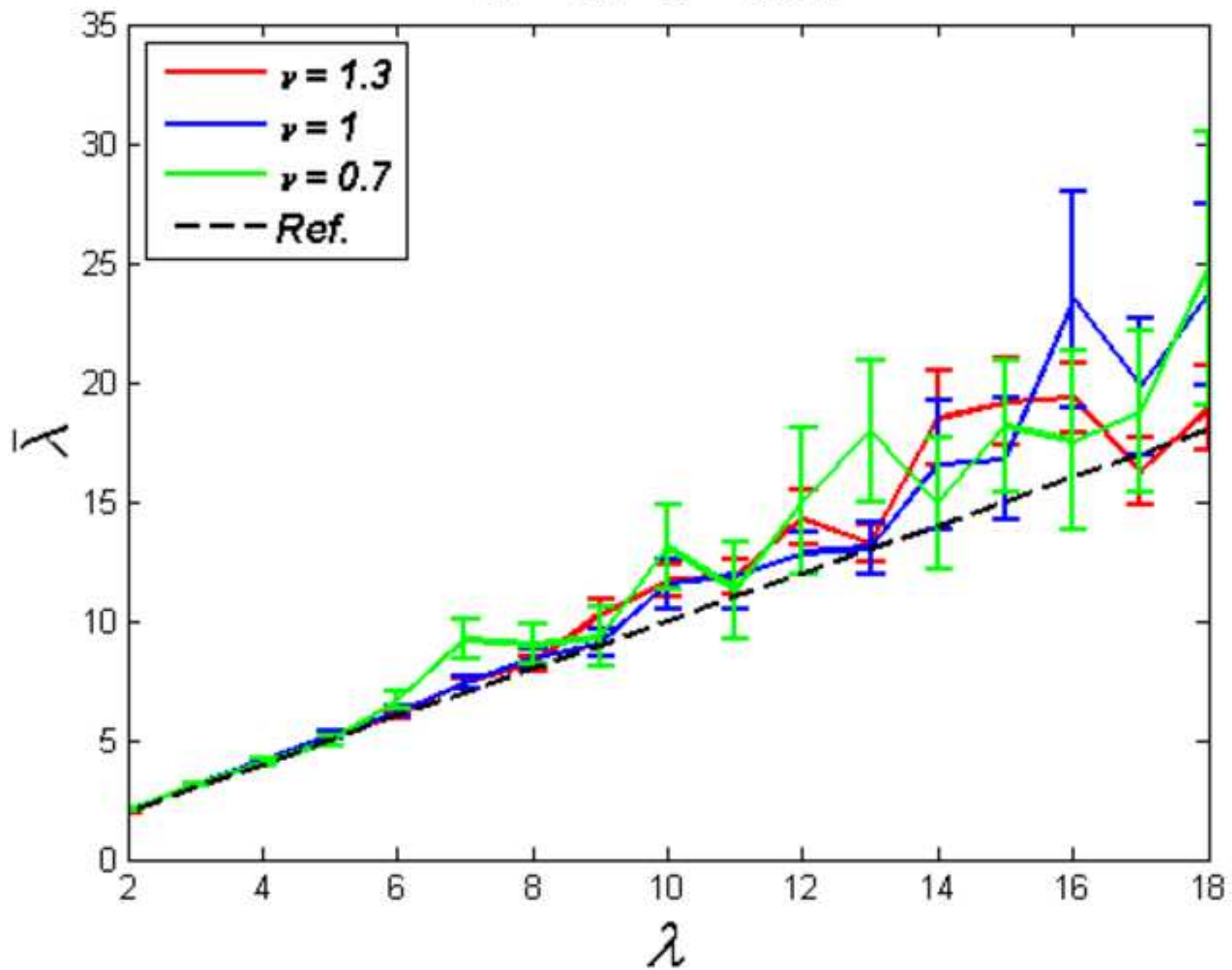


Figure 4d
[Click here to download high resolution image](#)

$N = 100$ $n = 1000$

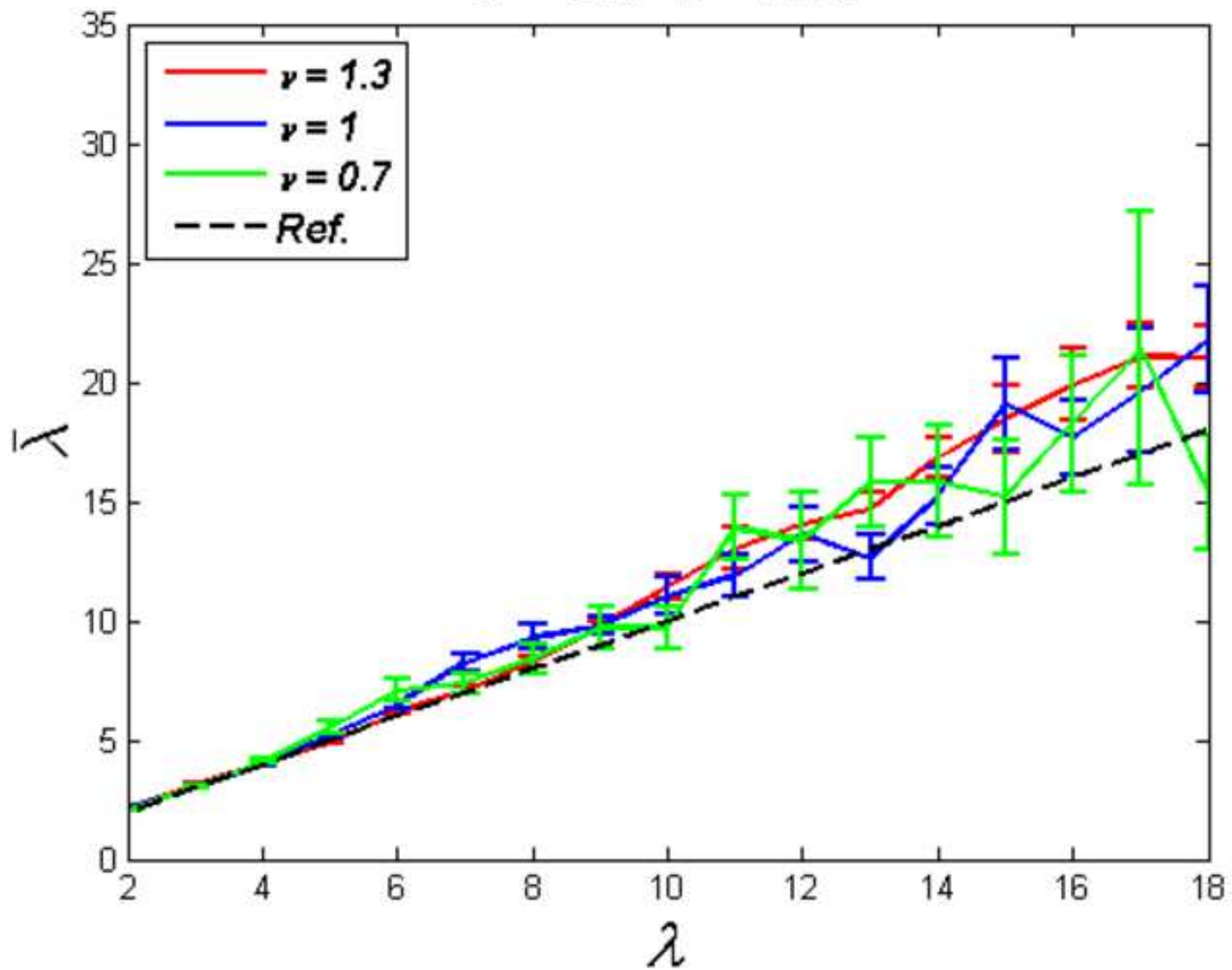


Figure 4e
[Click here to download high resolution image](#)

$N = 50$ $n = 5000$

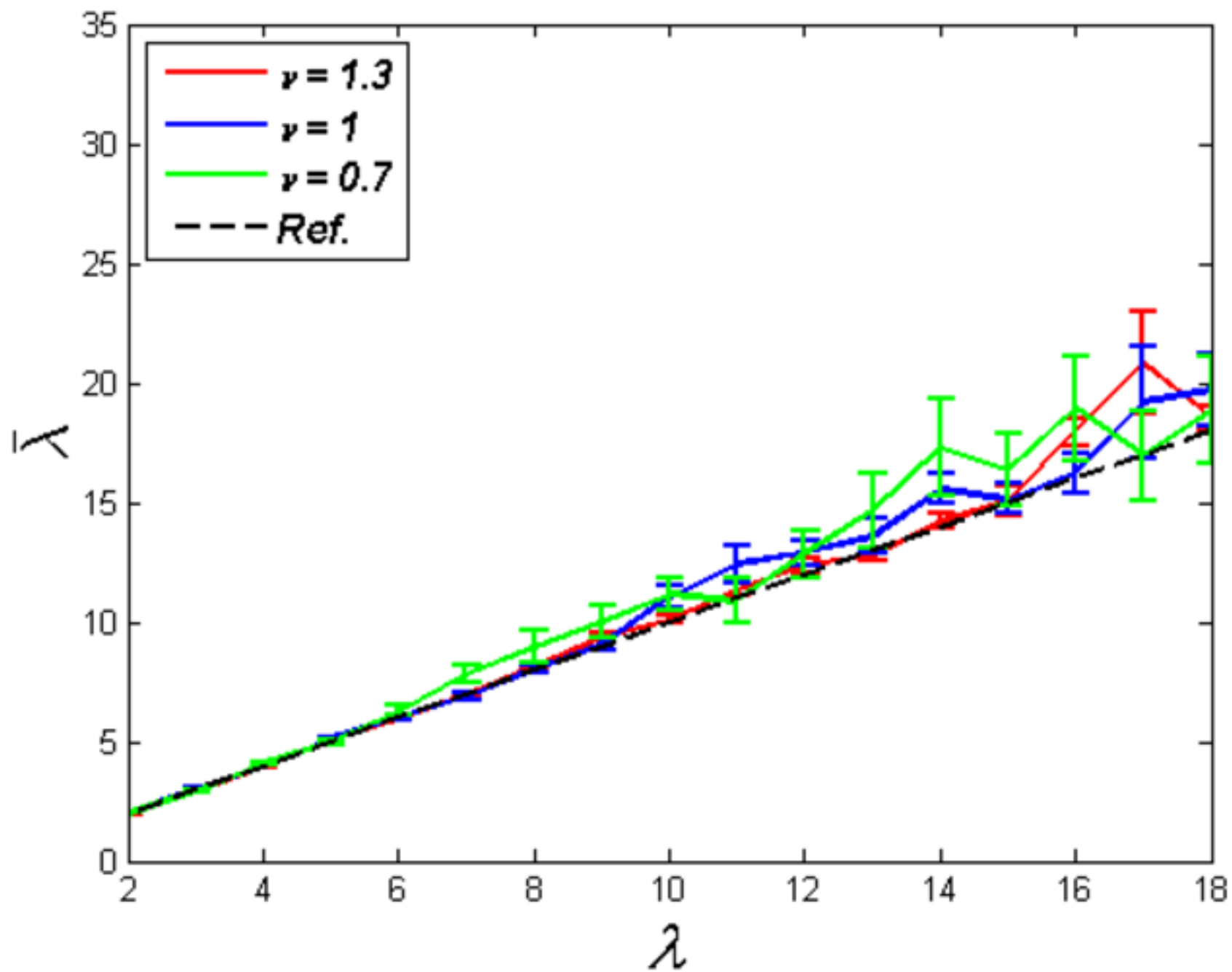


Figure 4f
[Click here to download high resolution image](#)

$N = 100$ $n = 5000$

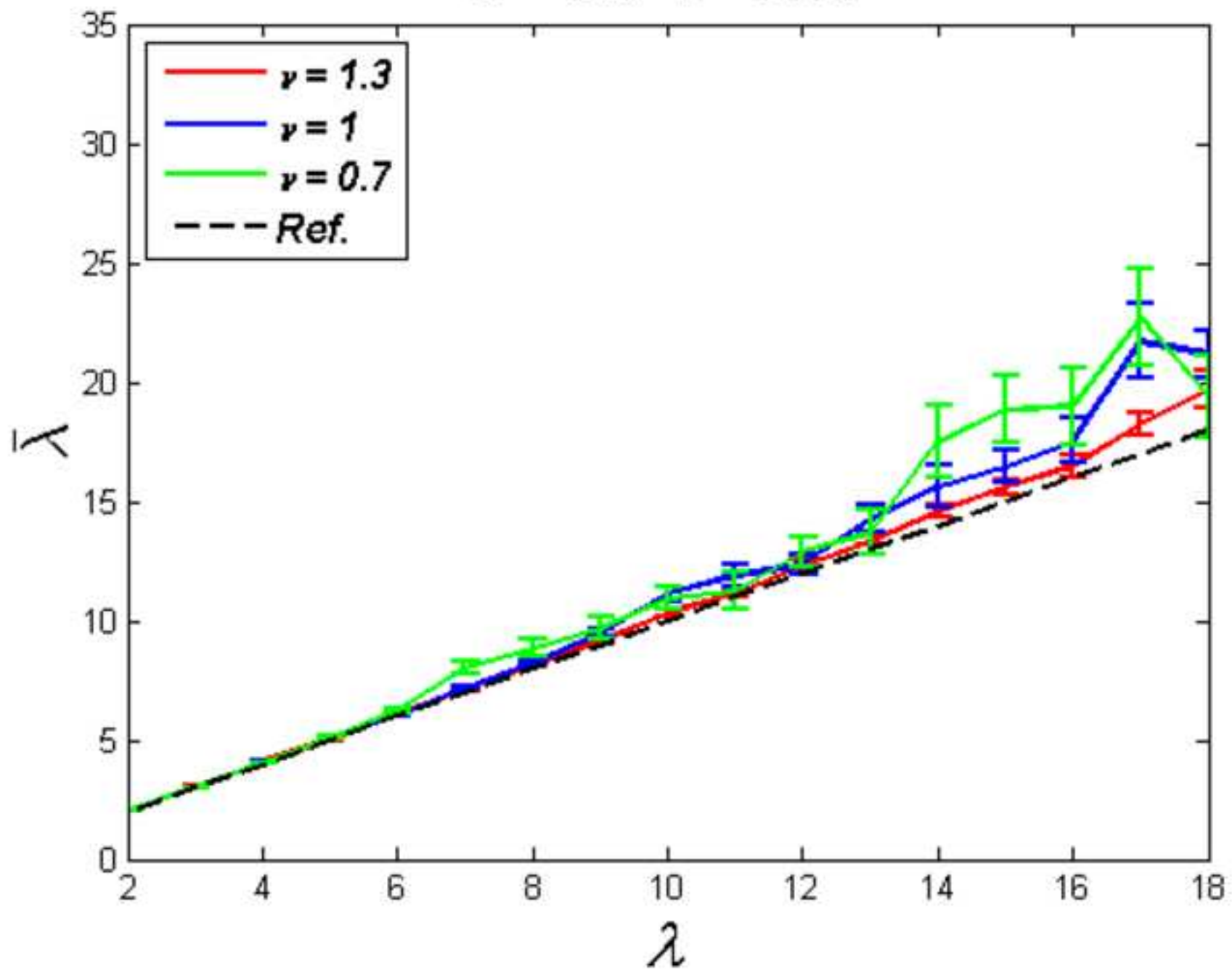


Figure 5a
[Click here to download high resolution image](#)

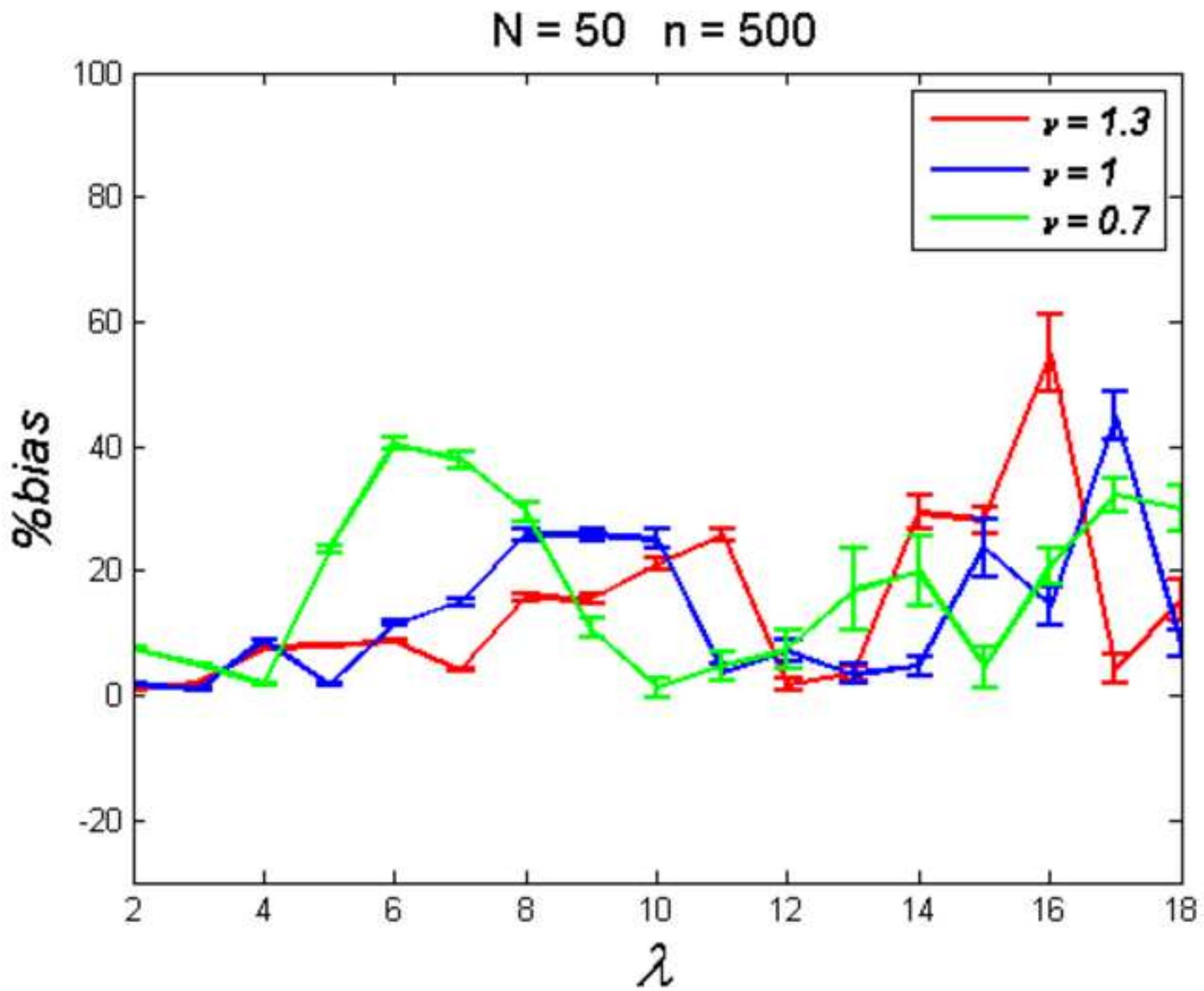


Figure 5b
[Click here to download high resolution image](#)

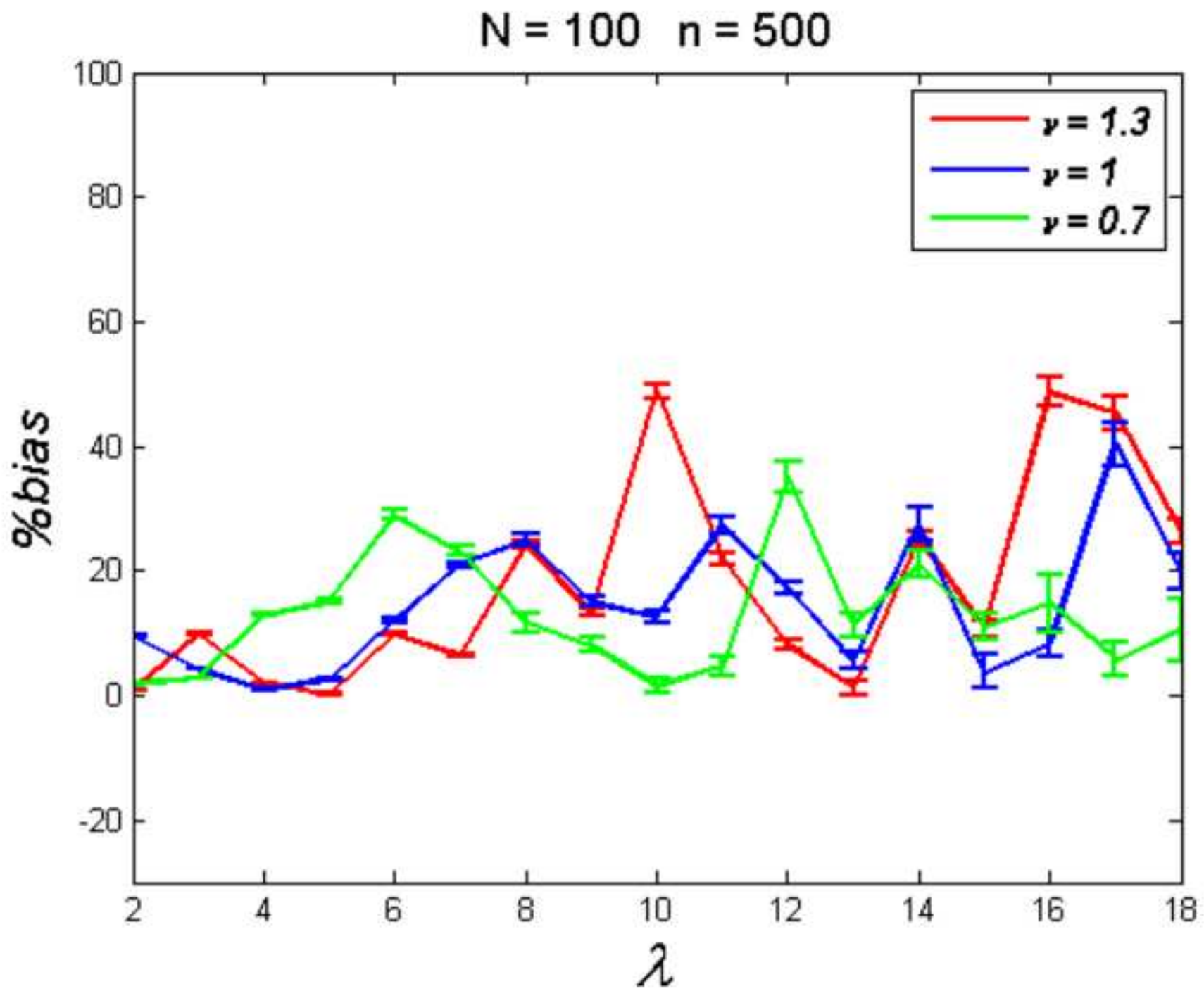


Figure 5c
[Click here to download high resolution image](#)

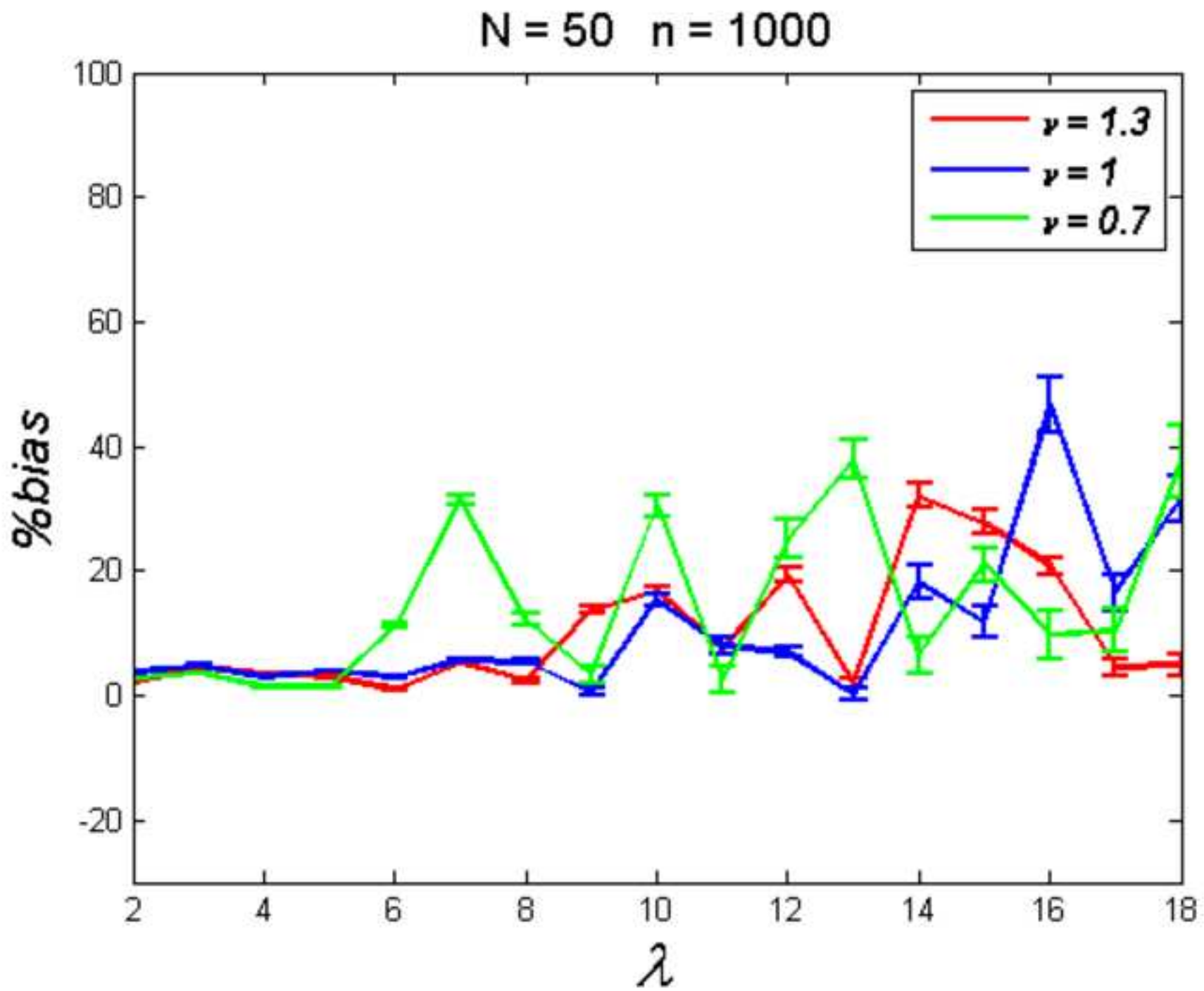


Figure 5d

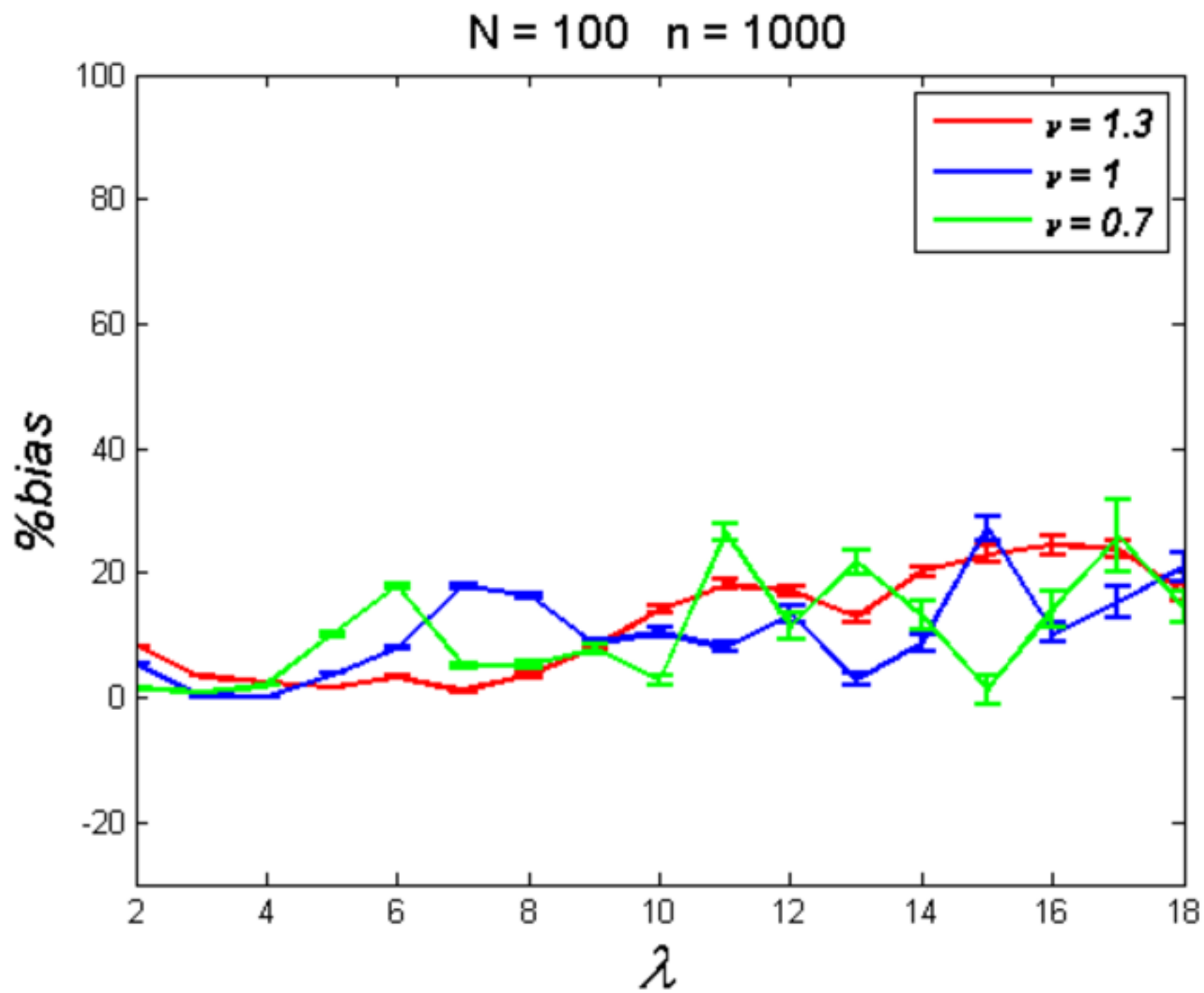
[Click here to download high resolution image](#)

Figure 5e
[Click here to download high resolution image](#)

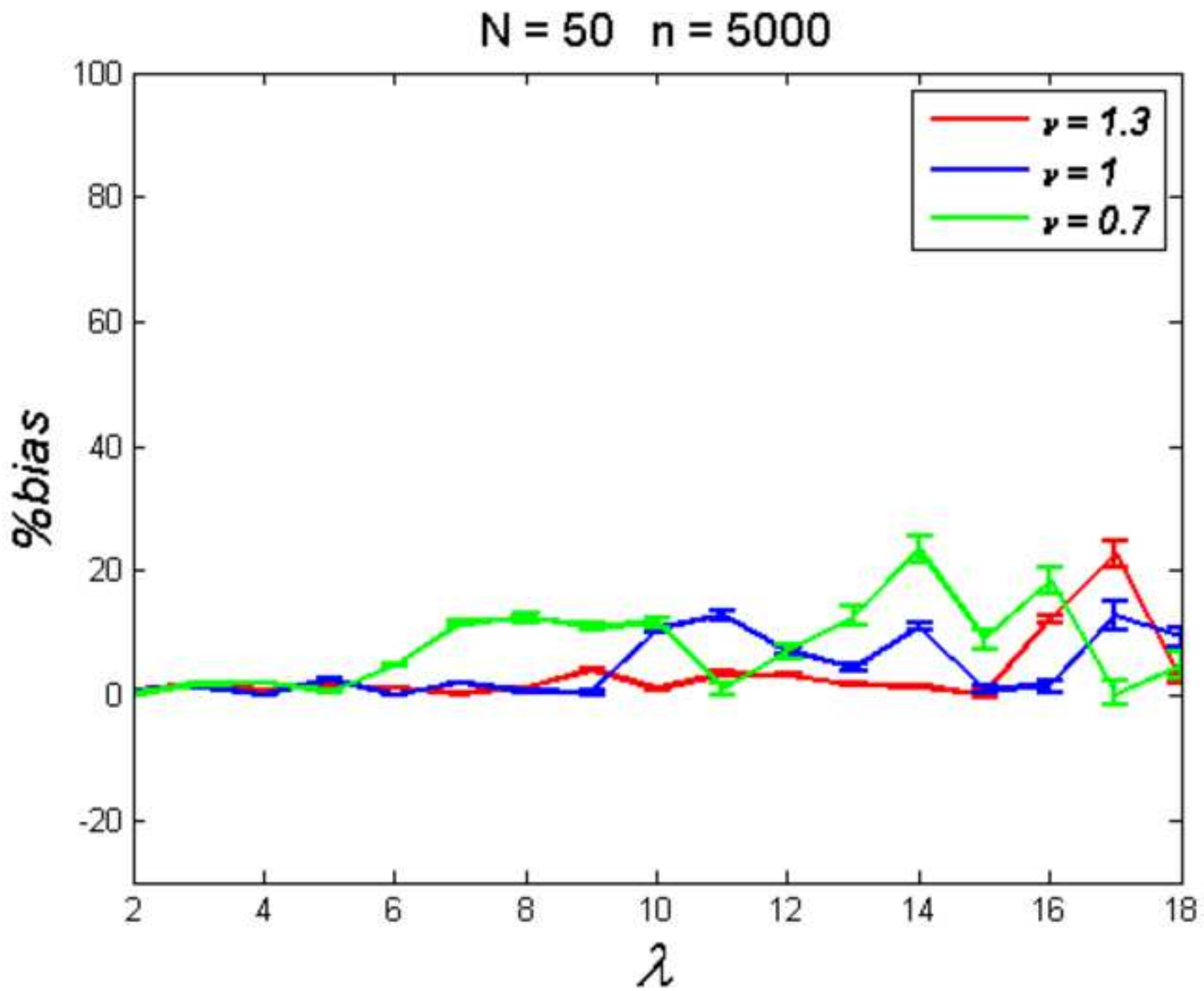


Figure 5f
[Click here to download high resolution image](#)

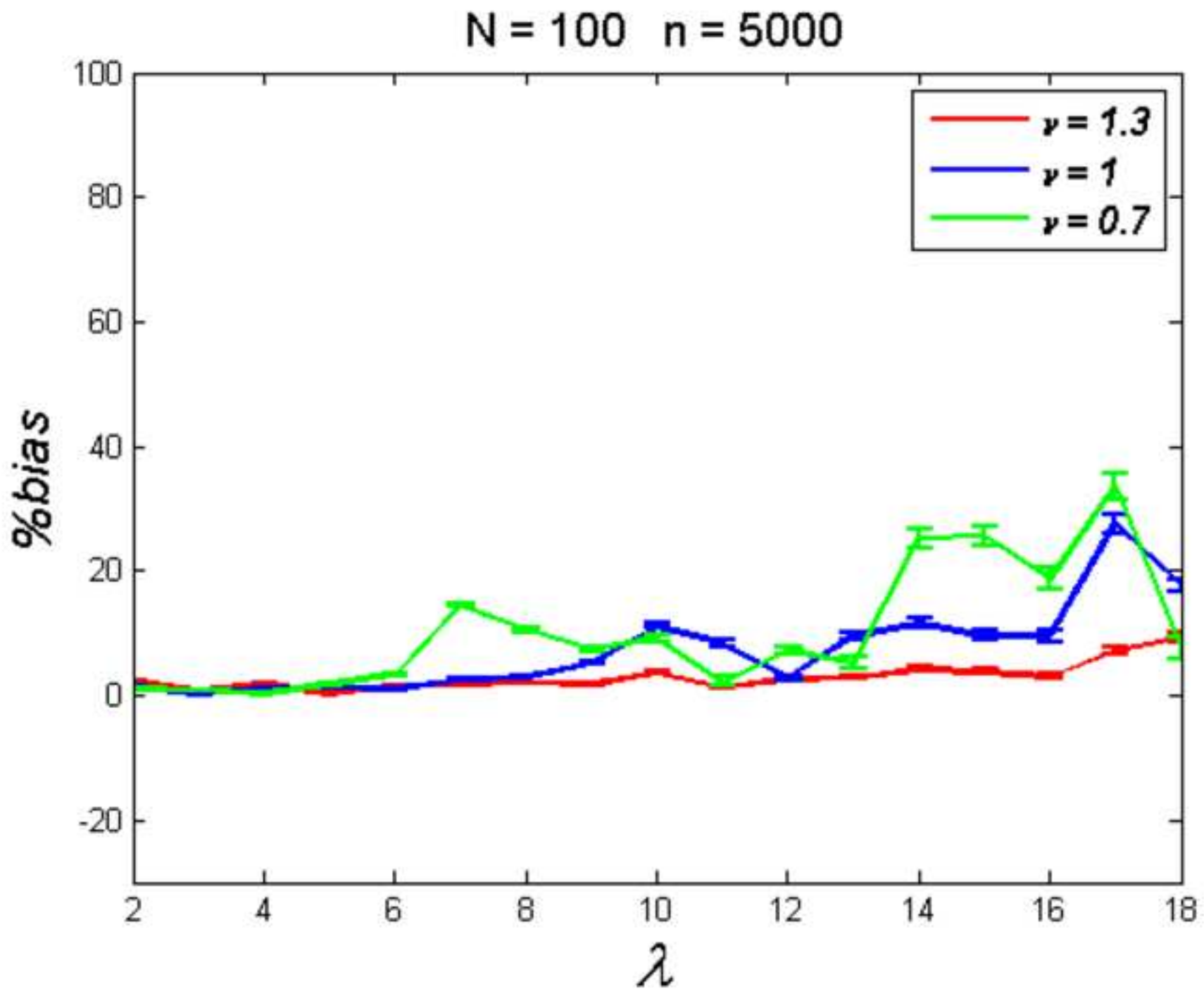


Figure 6
[Click here to download high resolution image](#)

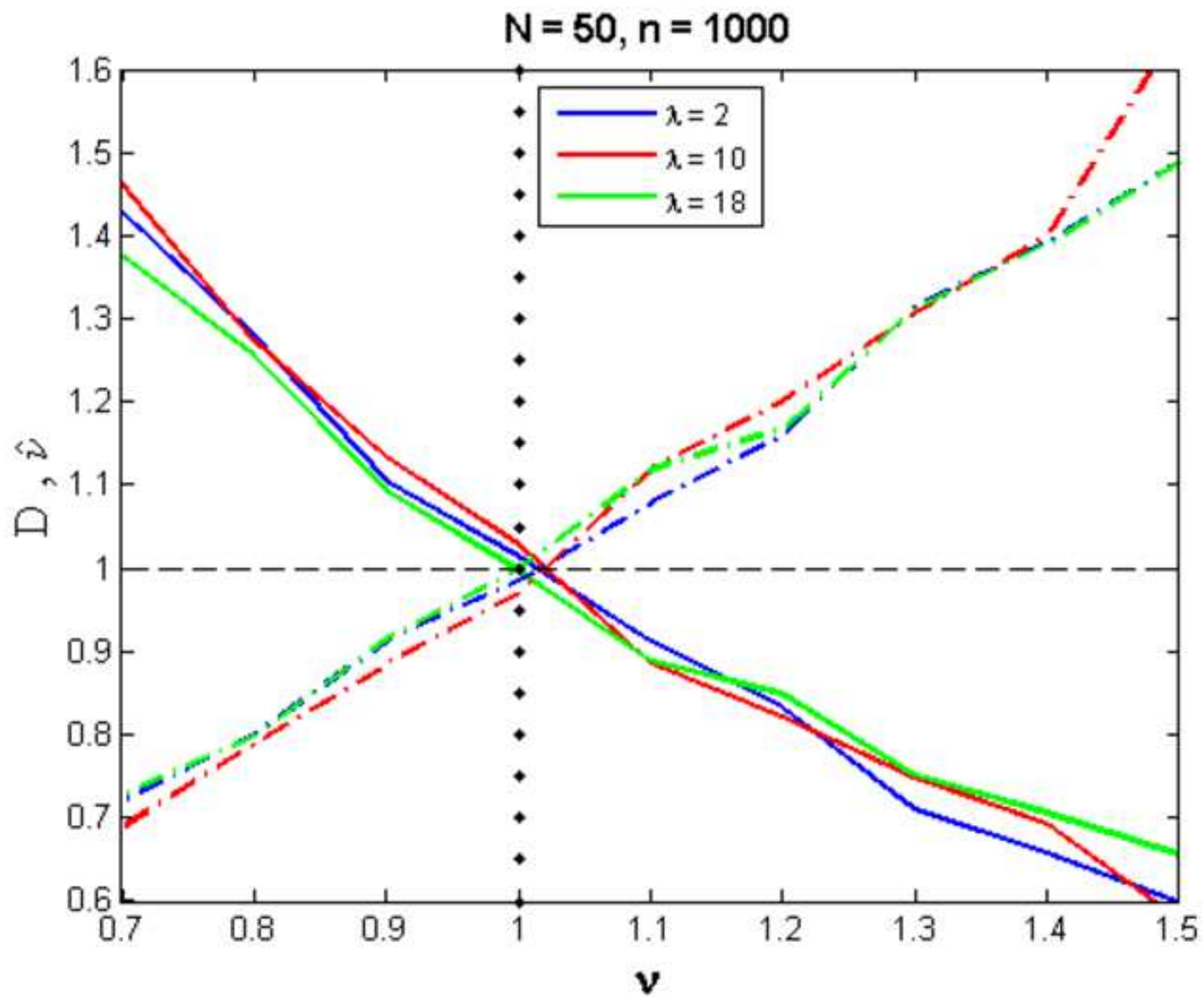


Figure 7a
[Click here to download high resolution image](#)

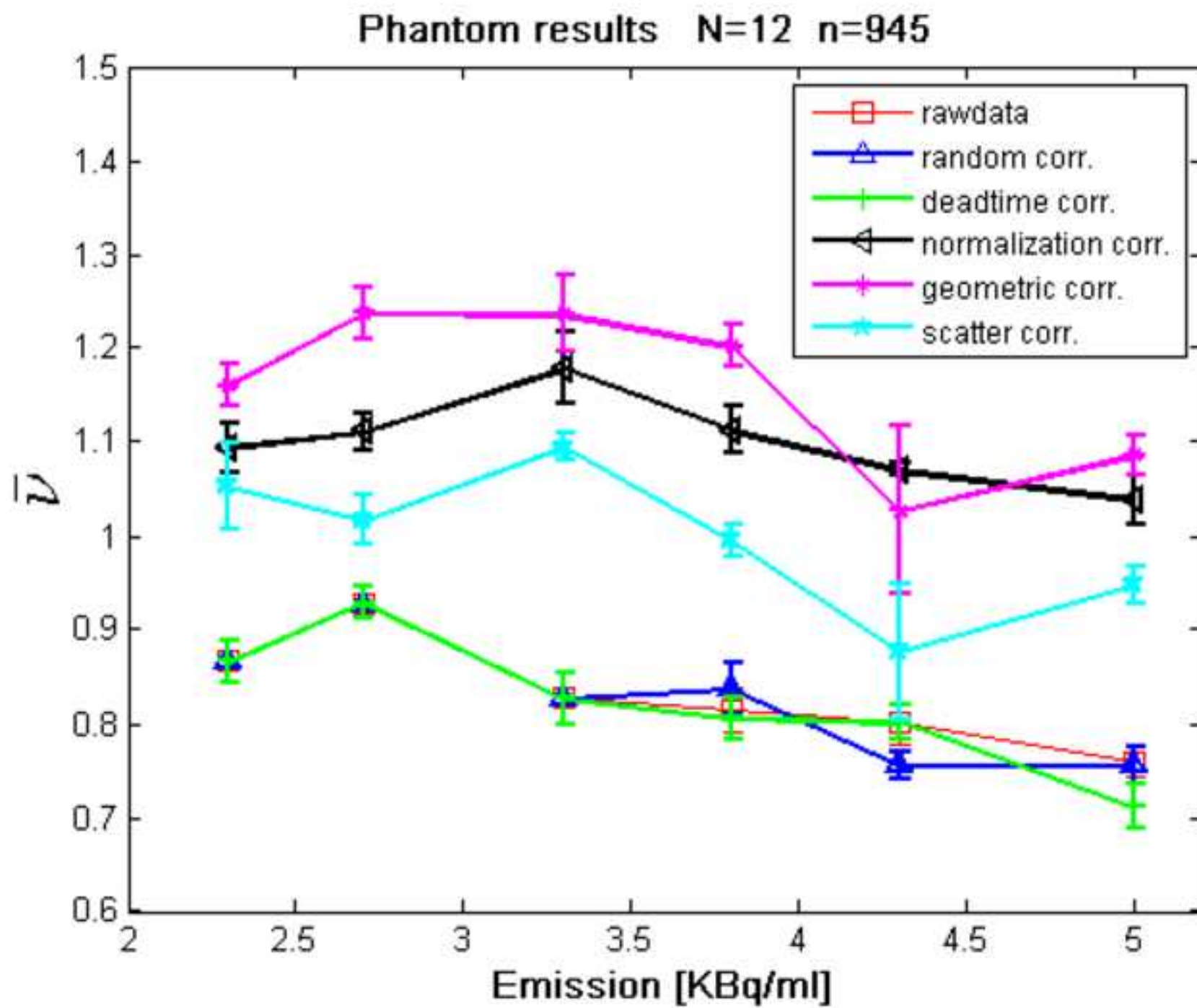


Figure 7b
[Click here to download high resolution image](#)

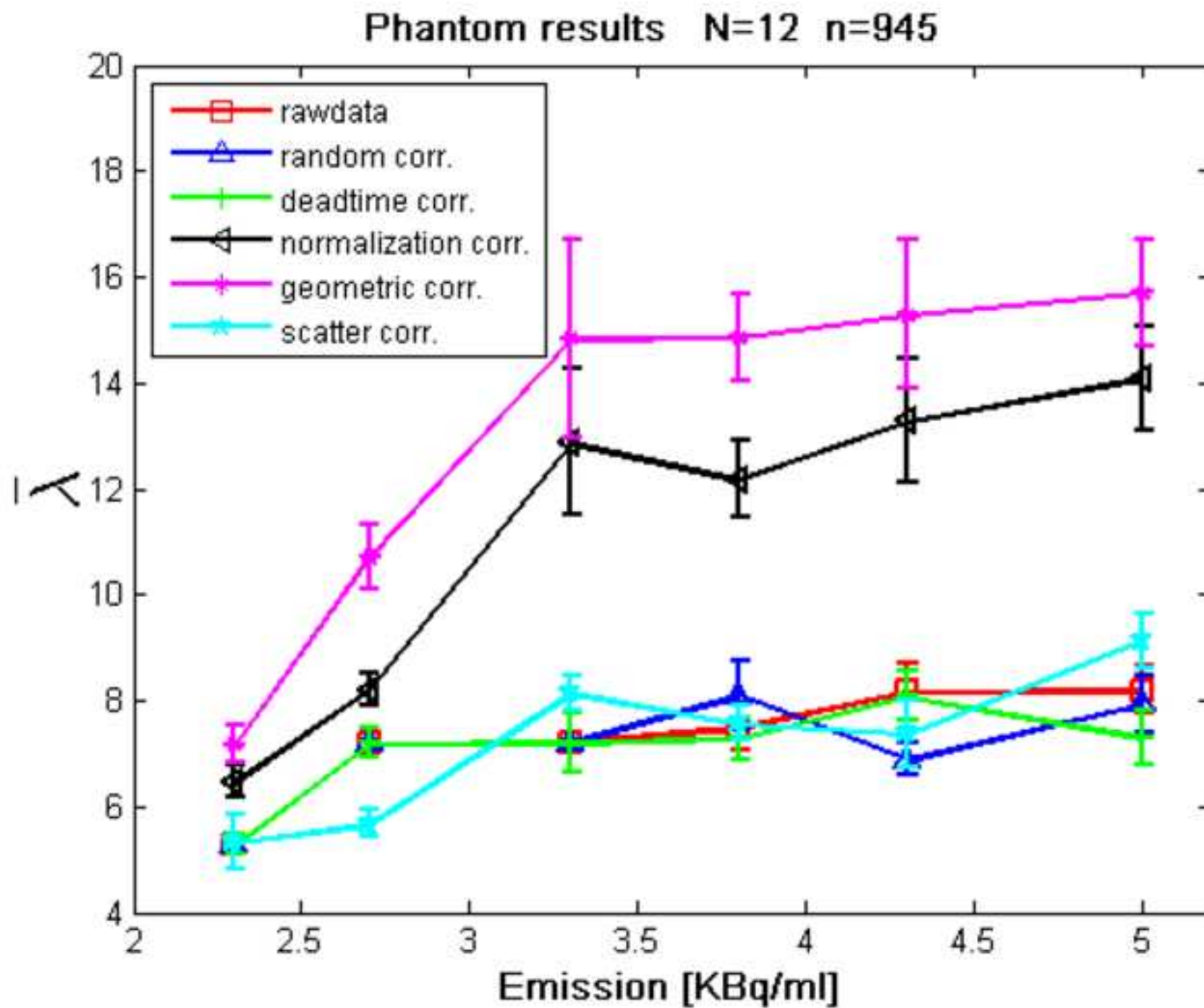


Figure 7c
[Click here to download high resolution image](#)

