Neural Network prediction of Riemann zeta zeros

O. Shanker *

Abstract

The theory underlying the location of zeroes of the Riemann Zeta Function is one of the key unsolved problems. Researchers are making extensive numerical studies to complement the theoretical work. The evaluation of the zeta function at large heights involves time consuming calculations. It would be helpful if one could have good predictions for the possible locations of the zeros, since one could then reduce the number of function evaluations required for locating the zeros. In this work we apply neural network regression as a tool to aid the empirical studies of the locations of the zeros. We use values evaluated at the Gram points as the input feature set for the predictions. The range of t studied is 267653395648.87 to 267653398472.51.

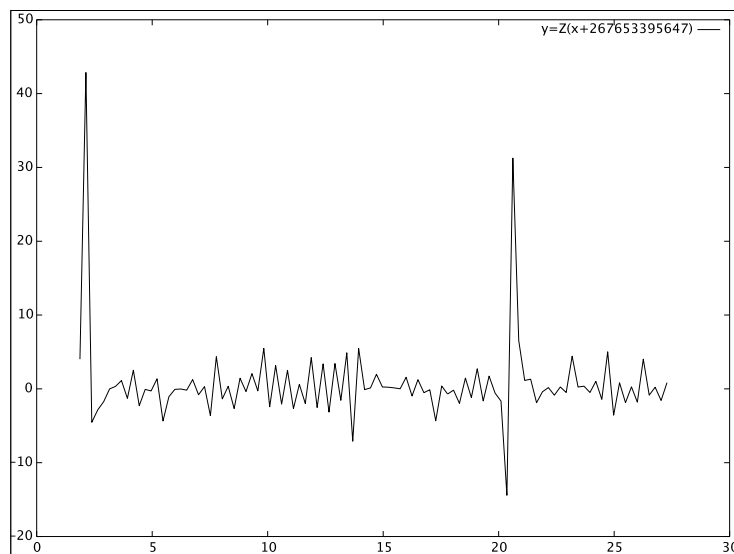*Mountain View, CA 94041, U. S. A. Email: oshanker@gmail.com

Figure 1: Riemann-Siegal function $Z(g_n)$ evaluated at 500 Gram points. X-axis is $t - 267653395647$.

1   Introduction

In the 1850's Riemann came up with an intriguing hypothesis about the location of the roots of the Riemann zeta function, which challenges us to this day. There is an impressive body of empirical evidence for his hypothesis, but a formal proof has been elusive. The numerical studies have found regions where the hypothesis is almost violated, but no counter-example has been found. The numerical calculations are time-consuming, and would benefit from tools which give approximate locations for the zeros of the function. We apply neural network regression to aid the empirical studies of the locations of the zeros. The paper is organised as follows. Section 2 establishes the required notation for the Riemann Zeta Function and L-functions. We also review breifly the theoretical work related to the locations of the zeros to show the importance of the topic. Section 3 describes the numerical evaluation of the Riemann zeta function. In section 4 we present results on training a neural network to predict the zero locations. The feature set used for training the neural network is based on the values at Gram points. The range of t studied is 267653395648.87 to 267653398472.51, which encompasses the range of zeros from 10^{12} to $10^{12} + 10^4$. Section 5 gives a brief summary of the results.

2 Theory of zero distributions of Generalised zeta functions

In this section we establish the required notation for the Riemann Zeta Function and L-functions. We also discuss the theory of the zero distributions, which gives the subject its importance.

The Riemann Zeta function is defined for $\text{Re}(s) > 1$ by

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \prod_{p \in primes} \left(1 - p^{-s}\right)^{-1}. \tag{1}$$

Eq. (1) converges for $\text{Re}(s) > 1$. It was shown by Riemann [1, 2, 3, 4] that $\zeta(s)$ has a continuation to the complex plane and satisfies a functional equation

$$\xi(s) := \pi^{-s/2} \Gamma(s/2) \zeta(s) = \xi(1 - s); \tag{2}$$

$\xi(s)$ is entire except for simple poles at $s = 0$ and 1. Riemann multiplied the definition by $s(s-1)$ to remove the poles. We write the zeroes of $\xi(s)$ as $1/2+i\gamma$. The Riemann Hypothesis asserts that γ is real for the non-trivial zeroes. We order the γs in increasing order, with

$$\ldots\ldots\gamma_{-1} < 0 < \gamma_1 \leq \gamma_2 \ldots. \tag{3}$$

Then $\gamma_j = -\gamma_{-j}$ for $j = 1, 2, \ldots$, and γ_1, γ_2, $\ldots$ are roughly 14.1347, 21.0220, ….

Asymptotically, for the Riemann zeta function the mean number of zeros with height less than γ (the smoothed Riemann zeta staircase) is [4]

$$< \mathcal{N}_{\mathcal{R}}(\gamma) >= (\gamma/2\pi)(ln(\gamma/2\pi) - 1) - \frac{7}{8}. \tag{4}$$

Thus, the mean spacing of the zeros at height γ is $2\pi(\ln(\gamma/2\pi))^{-1}$. For the range of t values studied by us this spacing is essentially constant at 0.2567.

The study of the zeroes of the Riemann zeta function and Generalised Zeta functions is of interest to mathematicians and physicists. Mathematicians study the spacings because of its applications to analytic number theory, while physicists study it because of its relation to the theory of the spectra of random matrix theories (RMT) and the spectra of classically chaotic quantum systems.

Odlyzko [5, 6] has made extensive numerical studies of the zeroes of the Riemann zeta function and their local spacings. He also studied their relation to the random matrix models of physics. Wigner [7] suggested that the resonance lines of a heavy nucleus might be modeled by the spectrum of a large random matrix. Gaudin [9] and Gaudin-Mehta [8] gave results for the local (scaled) spacing distributions between the eigenvalues of typical members of the ensembles as $N \to \infty$, based on their study of orthogonal polynomials. Later Dyson [10] introduced the closely related circular ensembles.

Odlyzko confirmed numerically that the local spacings of the zeroes of the Riemann Zeta function obey the laws for the (scaled) spacings between the

eigenvalues of a typical large unitary matrix. That is, they obey the laws of the Gaussian Unitary Ensemble (GUE) [7, 8, 9, 10]. Katz and Sarnak [12] state that at the phenomenological level this may be the most striking discovery about zeta since Riemann. Odlyzko's computations thus verified the discoveries and conjectures of Montgomery [13, 14, 15].

Further evidence for the connection between random matrices and generalised zeta functions comes from calculations of the zero correlation functions [16, 17, 18, 19], and the study of the low-lying zeros of families of L-functions [20]. Extensive numerical computations [5, 11] have strengthened the connection.

Several authors [21, 22, 23, 24, 25] have studied the moments of the Riemann zeta function and families of L-functions, and the relation to characteristic polynomials of random matrices. The autocorrelation functions were studied in [26, 27]. The relation of the Riemann zeta function to probability laws was studied in [28]. The author of this work studied the distributions of the zero spacings using Rescaled Range Analysis [29].

It has been shown that the long-range statistics of the zeroes of the Riemann zeta function are better described in terms of primes than by the GUE RMT. Berry [30, 31, 32, 33] has related this to a study of the semiclassical behaviour of classically chaotic physical systems. The primitive closed orbits of the physical system are analogous to the primes p. The analogy comes from formulae that connect zeros of the zeta function and prime numbers [34, 35, 36]

Quantum chaology is defined by Berry [32] as the study of semiclassical, but non-classical, behaviour characteristic of systems whose classical motion exhibits chaos. By semiclassical one means the limit as Planck's constant $\hbar \to 0$. The distribution of eigenvalues of quantum chaotic systems shows universality. The universality class depends on the symmetries of the system's Hamiltonian. For systems without time-reversal invariance the distribution of eigenvalue spacings approaches that of the GUE. The connection between quantum chaology and the Riemann zeta function comes about because the Riemann Hypothesis would follow if the imaginary parts γ_j of the non-trivial zeros of the Riemann zeta function are eigenvalues of a self-adjoint operator.

The remarkable properties of the Riemann Zeta Function can be generalised to a host of other zeta and L-functions. The simplest of the generalisations are for the Dirichlet L-functions $L(s, \chi)$ defined as follows: $q \geq 1$ is an integer and χ is a primitive character of the Abelian group formed by all integers smaller than and relatively prime to q. χ is extended to all integer values by making it periodic, and $\chi(m) = 0$ if m and q have a common factor. Then

$$L(s, \chi) = \sum_{n=1}^{\infty} \chi(n) n^{-s} = \prod_{p} \left(1 - \chi(p) p^{-s}\right)^{-1}. \qquad (5)$$

The analogue of the functional equation Eq. (2) is known for the generalised zeta functions, and they also seem to satisfy the generalised Riemann Hypothesis. q is called the conductor of the L-function. In this work we perform studies on
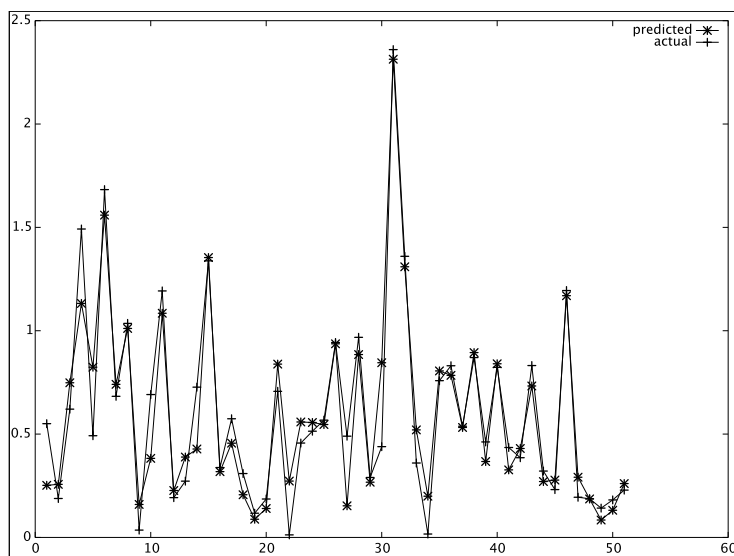
Figure 2: Comparision of neural network prediction vs actual value for the distance from a Gram point to the next zero. The x axis shows 50 Gram points beginning at index 1000000008584.

the Riemann zeta function. It would be worthwhile to extend the study to the other zeta functions.

The next section gives the details of the numerical calculations.

3   Empirical Calculations

In this section we discuss the details of the numerical work. The numerical analysis takes advantage of the functional equation Eq. (2). One defines

$$\theta(t) = arg(\pi^{it/2}\Gamma(\frac{1}{4} + \frac{it}{2})), \tag{6}$$

where the argument is defined by continuous variation of t starting with the value 0 at $t = 0$. For large t θ has the aymptotic expansion

$$\theta(t) = \frac{t}{2}\ln(\frac{t}{2\pi}) - \frac{t}{2} - \frac{\pi}{8} + \frac{1}{48t} - \frac{1}{5760t^3}. \tag{7}$$

A consequence of the zeta functional equation is that the function $Z(t) = exp(i\theta(t))\gamma(1/2 + it)$, known as the Riemann-Siegel Z-function, is real valued for real t. Moreover we have $|Z(t)| = |\zeta(1/2 + it)|$. Thus the zeros of $Z(t)$ are the imaginary part of the zeros of $\zeta(s)$ which lie on the critical line. We are led to finding the change of sign of a real valued function to find zeros on the critical line. This is a very convenient property in the numerical verification

721

of the Riemann Hypothesis. Another very helpful property is that many of the zeros are separated by the "Gram points". When $t \geq 7$, the θ function Eq.(6) is monotonic increasing. For $n \geq 1$, the $n - th$ Gram point g_n is defined as the unique solution > 7 to $\theta(g_n) = n\pi$. The Gram points are as dense as the zeros of $\zeta(s)$ but are much more regularly distributed. Their locations can be found without any evaluations of the Riemann-Siegal series Eq.(8). Grams law is the empirical observation that $Z(t)$ usually changes its sign in each Gram interval $G_n = [g_n, g_{n+1})$. This law fails infinitely often, but it is true in a large proportion of cases. The average value of $Z(g_n)$ is 2 for even n and -2 for odd n [3], and hence $Z(g_n)$ undergoes an infinite number of sign changes. Figure 1 shows the Riemann-Siegal function $Z(g_n)$ evaluated at 500 Gram points starting at $n = 99999999999$. Given the desirable properties of the Gram points, it seems natural to use the values at these points as the feature set for the neural network regression.

The Riemann-Siegel Z-function is evaluated using the Riemann-Siegal series

$$Z(t) = 2 \sum_{n=1}^{m} \frac{\cos(\theta(t) - t \ln(n))}{\sqrt{n}} + R(t), \tag{8}$$

where m is the integer part of $\sqrt{t/(2\pi)}$, and $R(t)$ is a small remainder term which can be evaluated to the desired level of accuracy. The most important source for loss of accuracy at large heights is the cancellation between large numbers that occur in the arguments of the cos terms in Eq. (8). We use a high precision module to evaluate the arguments. The rest of the calculation is done using regular double precision accuracy. The range of t studied is 267653395648.87 to 267653398472.51, which encompasses the range of zeros from 10^{12} to $10^{12} + 10^4$. Odlyzko [5] has published very accurate values for the zeros in this range. By comparing our calculations with the the calculations of Odlyzko, we estimate that our accuracy for the Z function evaluation is better than 10^{-8}.

In the next section we discuss the application of neural network regression to aid the location of the zeros.

4   Neural Network Regression

This section describes the feature set used in training our neural networks, and the results of the training. Neural Network Regression (NNR) models have been applied for forecasting and are included in state of the art techniques in forecasting methods [37, 38, 39]. Cybenkos [40] and Hornik et al. [41] demonstrated that specific NNR models can approximate any continuous function if their hidden layer is sufficiently large. Furthermore, NNR models are equivalent to several nonlinear nonparametric models, i.e. models where no decisive assumption about the generating process must be made in advance [42]. Kouam et al. [43] have shown that most forecasting models (ARMA models, bilinear models, autoregressive models with thresholds, non- parametric models with kernel regression, etc.) are embedded in NNR. The advantage of NNR models can be summarised as follows: if, in practice, the best model for a given
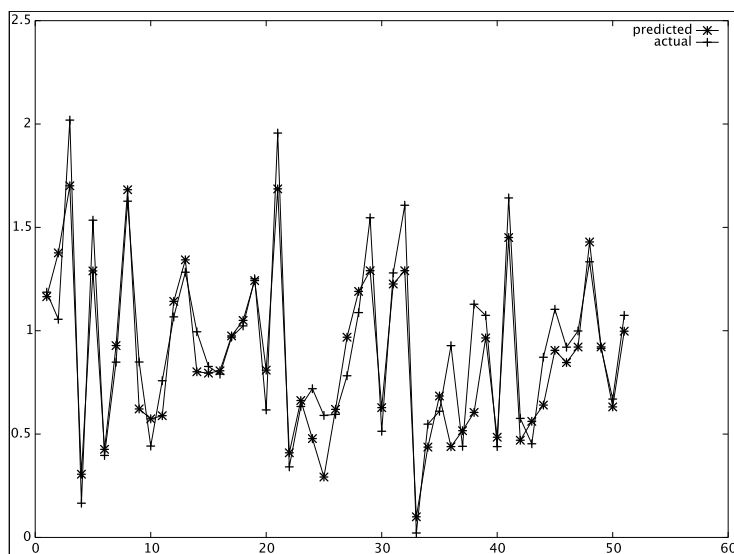
Figure 3: Comparision of neural network prediction vs actual value for the smallest zero difference in the Gram interval (the left hand zero lies in the Gram interval). The x axis shows 50 Gram points beginning at index 1000000008584.

problem cannot be determined, it is a good idea to use a modelling strategy which is a generalisation of a large number of models, rather than to impose a priori a given model specification. This is the reason for the interest in NNR applications [44, 45, 46, 47, 48, 49].

As explained in Section 3, we use values defined at the Gram points for the input feature set provided to the neural network. This is because the positions of the Gram points can be determined to the desired accuracy without extensive costly calculations. The features we use are the values of the Riemann-Segal Z-function, the first ten terms in the Riemann-Segal series Eq. (8), and nine terms with the cos function replaced by the respective sin function, for a pair of consecutive Gram points. We drop the first sin term because it is always zero at Gram points. Thus, the size of the input feature set is 40. We use a two layer neural network with 200 neurons in the hidden layer.

We train the neural network to separately predict two quantities. The first is the distance from a Gram point to the next zero. The second is the smallest zero difference in the Gram interval with the left hand zero lying in the Gram interval. The reason for choosing these output functions is that close pairs of zeros are interesting. They correspond to cases for which the RH is nearly false. Calculating the Riemann zeta function in zones where two zeros are very close is a stringent test of the Riemann Hypothesis (often described as Lehmer's phenomenon since Lehmer was the first to observe such situations). We use the scaled difference:

$$\delta_j = (\gamma_{j+1} - \gamma_j) \ln(\gamma_j/2\pi)/(2\pi). \tag{9}$$

723

Next zero	Mean	Std.Deviation
Training set		
Acutal	0.579	0.41
Predicted	0.579	0.37
Validation set		
Acutal	0.599	0.42
Predicted	0.590	0.38
Test set		
Acutal	0.576	0.43
Predicted	0.581	0.39

Table 1: Prediction of the distance from a Gram point to the next zero.

Smallest zero difference	Mean	Std.Deviation
Training set		
Acutal	0.965	0.41
Predicted	0.926	0.39
Validation set		
Acutal	0.959	0.41
Predicted	0.926	0.39
Test set		
Acutal	0.972	0.43
Predicted	0.934	0.42

Table 2: Prediction of the smallest zero difference in the Gram interval with the left hand zero lying in the Gram interval.

We divide the 10000 data points into three sets, of sizes 9000, 500 and 500 respectively. The first set of 9000 points is used for training the neural network. The second set of 500 points is used as a validation set during the training. Finally, the last set of 500 points is used to check how well the neural network has learned to predict the output function.

Figure 2 shows the comparision of the neural network prediction vs actual value for the distance from a Gram point to the next zero for 50 Gram points beginning at index 1000000008584. Table 1 presents summary results for the complete data set. The table shows the the mean value and standard deviation for the actual data set and also for the predicted values. Figure 3 shows the comparision of the neural network prediction vs actual value for the smallest zero difference in the Gram interval with the left hand zero lying in the Gram interval, for the same 50 Gram points. Table 2 presents summary results for the complete data set. The table shows the the mean value and standard deviation for the actual data set and also for the predicted values.

In both cases, we see that the neural network gives fairly good predictions. It follows the peaks and dips of the output function faithfully. It is successful in reproducing the mean value as well as the variation of the output it is attempting to predict. Given this encouraging result, it seems worthwhile to consider using more features in the input layer and more neurons in the hidden layer to increase the fidelity of the predictions.

In the next section we give a brief summary of the results.

5   Conclusions

We studied the Riemann zeta function for the range of t from 267653395648.87 to 267653398472.51. We trained neural networks to predict the distance from a Gram point to the next zero, and the smallest zero difference in the Gram interval with the left hand zero lying in the Gram interval. In both cases, the neural network gives fairly good predictions. It follows the peaks and dips of the output function faithfully. It is successful in reproducing the mean value as well as the variation of the output it is attempting to predict. Given this encouraging result, it seems worthwhile to consider using more features in the input layer and more neurons in the hidden layer to increase the fidelity of the predictions. The use of neural networks should also be extended to the Generalised zeta functions.

References

[1] B. Riemann, "Über die Anzahl der Primzahlen uter Einer Gegebenen Gröbe," *Montasb. der Berliner Akad.*, (1858), 671-680.

[2] B. Riemann, "Gesammelte Werke", Teubner, Leipzig, (1892).

[3] E. Titchmarsh, "The Theory of the Riemann Zeta Function," Oxford University Press, Second Edition, (1986).

[4] H. M. Edwards, "Riemann's Zeta Function," Academic Press, (1974).

[5] A. Odlyzko, "The 10^{20}-th Zero of the Riemann Zeta Function and 70 Million of its Neighbors," (preprint), A.T.T., (1989).

[6] A. Odlyzko, "Dynamical, Spectral, and Arithmetic Zeta Functions", Amer. Math. Soc., Contemporary Math. series, **290**, 139-144, (2001).

[7] E. Wigner, "Random Matrices in Physics," *Siam Review*, **9**, 1-23, (1967).

[8] M. Gaudin, M. Mehta, "On the Density of Eigenvalues of a Random Matrix," *Nucl. Phys.*, **18**, 420-427, (1960).

[9] M. Gaudin, "Sur la loi Limite de L'espacement de Valuers Propres D'une Matrics Aleatiore," *Nucl. Phys.*, **25**, 447-458, (1961).

[10] F. Dyson, "Statistical Theory of Energy Levels III," *J. Math. Phys.*, **3**, 166-175, (1962).

[11] M. Rubinstein, *Evidence for a Spectral Interpretation of Zeros of L-functions*, PhD thesis, Princeton University, 1998.

[12] Nicholas M. Katz, Peter Sarnak, "Zeroes of zeta functions and symmetry," *Bulletin of the AMS*, **36** , 1-26, (1999).

[13] H. Montgomery, "Topics in Multiplicative Number Theory," L.N.M., **227**, Springer, (1971).

[14] H. Montgomery, "The Pair Correlation of Zeroes of the Zeta Function," *Proc. Sym. Pure Math.*, **24**, *AMS*, 181-193, (1973).

[15] D. Goldston, H. Montgomery, "Pair Correlation of Zeros and Primes in Short Intervals," *Progress in Math.*, Vol. 70, Birkhauser, 183-203, (1987).

[16] D.A. Hejhal, On the triple correlation of zeros of the zeta function, *Inter. Math. Res. Notices*, **7**:293–302, 1994.

[17] Z. Rudnick and P. Sarnak, Principal L-functions and random matrix theory, *Duke Mathematical Journal*, **81**(2):269–322, 1996.

[18] E.B. Bogomolny and J.P. Keating, Random matrix theory and the Riemann zeros I: three- and four-point correlations, *Nonlinearity*, **8**:1115–1131, 1995.

[19] E.B. Bogomolny and J.P. Keating, Random matrix theory and the Riemann zeros II:n-point correlations, *Nonlinearity*, **9**:911–935, 1996.

[20] Nicholas M. Katz, Peter Sarnak, *Random Matrices, Frobenius Eigenvalues and Monodromy*, AMS, Providence, Rhode Island, 1999.

[21] J.P. Keating and N.C. Snaith, Random matrix theory and $\zeta(1/2 + it)$, *Commun. Math. Phys.*, **214**:57–89, 2000.

[22] J.P. Keating and N.C. Snaith, Random matrix theory and *L*-functions at $s = 1/2$, *Commun. Math. Phys*, **214**:91–110, 2000.

[23] J.B. Conrey and D.W. Farmer, Mean values of *L*-functions and symmetry, *Int. Math. Res. Notices*, **17**:883–908, 2000, arXiv:math.nt/9912107.

[24] C.P. Hughes, J.P. Keating, and N. O'Connell, Random matrix theory and the derivative of the Riemann zeta function, *Proc. R. Soc. Lond. A*, **456**:2611–2627, 2000.

[25] C.P. Hughes, J.P. Keating, and N. O'Connell, On the characteristic polynomial of a random unitary matrix, *Commun. Math. Phys.*, **220**(2):429–451, 2001.

[26] J.B. Conrey, D.W. Farmer, J.P. Keating, M.O. Rubinstein, and N.C. Snaith, Integral moments of zeta- and *L*-functions, *preprint*, 2002, arXiv:math.nt/0206018.

[27] J.B. Conrey, D.W. Farmer, J.P. Keating, M.O. Rubinstein, and N.C. Snaith, Autocorrelation of Random Matrix Polynomials *Commun. Math. Phys* , **237**:365-395, 2003

[28] Philippe Biane, Jim Pitman, and Marc Yor, "Probability Laws related to the Jacobi Theta and Riemann Zeta Functions, and Brownian Motion" *Bulletin of the AMS*, **38** , 435-465, (2001).

[29] O. Shanker, Generalised Zeta Functions and Self-Similarity of Zero Distributions, *J. Phys. A* **39**(2006), 13983-13997.

[30] M. V. Berry, "Semiclassical theory of spectral rigidity," *Proc. R. Soc.*, **A 400** , 229-251, (1985).

[31] M. V. Berry, "Riemann's zeta function: a model for quantum chaos?," *Quantum chaos and statistical nuclear physics (Springer Lecture Notes in Physics)*, **263** , 1-17, (1986).

[32] M. V. Berry, "Quantum Chaology," *Proc. R. Soc.* , **A 413** , 183-198, (1987).

[33] M. V. Berry, 'Number variance of the Riemann zeros'," *NonLinearity* , **1** ,399-407 , (1988).

[34] E. Landau, U:ber die Nullstellen der Zetafunktion, *Math. Ann. 71* (1911), 548-564.

[35] S. M. Gonek, A formula of Landau and mean values of $zeta(s)$, pp. 92-97 in *Topics in Analytic Number Theory*, S. W. Graham and J. D. Vaaler, eds., Univ. Texas Press, 1985.

[36] A. Odlyzko, "On the distribution of spacings between zeros of the zeta function", Math. Comp, **48**, 273-308, (1987).

[37] G. Zhang, B. E. Patuwo and M. Y. Hu, Forecasting with Artificial Neural Networks: The State of the Art, International Journal of Forecasting, **14**, 35-62, (1998).

[38] P. K. Simpson, Artificial Neural Systems - Foundations, Paradigms, Applications, and Implementations, Pergamon Press, New York, 1990.

[39] M. H. Hassoun, Fundamentals of Artificial Neural Networks, MIT Press, Cambridge, MA, 1995.

[40] G. Cybenko, "Approximation by Superposition of a Sigmoidal Function", Mathematical Control, Signals and Systems, **2**, 303-14, (1989).

[41] K. Hornik, M. Stinchcombe and H. White, "Multilayer Feedforward Networks Are Universal Approximators", Neural Networks, **2**, 359-66, (1989).

[42] B. Cheng and D. M. Titterington, Neural Networks : A Review from a Statistical Perspective, Statistical Science, **9**, 2-54, (1994).

[43] A. Kouam, F. Badran and S. Thiria, Approche Methodologique pour lEtude de la Prevision a lAide de Reseaux de Neurones, Actes de Neuro- Nimes, 1992.

[44] R. Trippi and E. Turban [eds.], Neural Networks in Finance and Investing - Using Artificial Intelligence to Improve Real-Word Performance, Probus Publishing Company, Chicago, 1993.

[45] G. J.Deboeck, Trading on the Edge - Neural, Genetic and Fuzzy Systems for Chaotic Financial Markets, John Wiley and Sons, New York, 1994.

[46] A. N.Refenes, Neural Networks in the Capital Markets, John Wiley and Sons, Chichester, 1995.

[47] H. Rehkugler and H. G. Zimmermann [eds.], Neuronale Netze in der Okonomie - Grundlagen und finanzwirtschaftliche Anwendungen, Verlag Franz Vahlen, Munchen, 1994.

[48] C. Dunis, The Economic Value of Neural Network Systems for Exchange Rate Forecasting, Neural Network World, **1**, 43-55, 1996.

[49] C. Dunis and X. Huang, Forecasting and Trading Currency Volatility: An Application of Recurrent Neural Regression and Model Combination, Liverpool Business School Working Paper, www.cibef.com, (2001).