risks

Article

Correlation Pitfalls with ChatGPT: Would You Fall for Them?

Marius Hofert

Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China; mhofert@hku.hk

Abstract: This paper presents an intellectual exchange with ChatGPT, an artificial intelligence chatbot, about correlation pitfalls in risk management. The exchange takes place in the form of a conversation that provides ChatGPT with context. The purpose of this conversation is to evaluate ChatGPT's understanding of correlation pitfalls, to offer readers an engaging alternative for learning about them, but also to identify related risks. Our findings indicate that ChatGPT possesses solid knowledge of basic and mostly non-technical aspects of the topic, but falls short in terms of the mathematical rigor needed to avoid certain pitfalls or completely comprehend the underlying concepts. Nonetheless, we suggest ways in which ChatGPT can be utilized to enhance one's own learning process.

Keywords: ChatGPT; correlation; dependence; pitfalls

MSC: 68T50

JEL Classification: C45; G32; D83

check for updates

Citation: Hofert, Marius. 2023. Correlation Pitfalls with ChatGPT: Would You Fall for Them? *Risks* 11: 115. https://doi.org/10.3390/risks11070115

Academic Editor: Mogens Steffensen

Received: 16 May 2023
Revised: 16 June 2023
Accepted: 19 June 2023
Published: 21 June 2023

1. Introduction

ChatGPT is a state-of-the-art artificial intelligence language model developed by research laboratory OpenAI. It was released to the public as a prototype on 30 November 2022, and has already gained significant attention for its capabilities (but also failures). ChatGPT utilizes specific *transformer (T)* neural networks which have been *pre-trained (P)* on vast amounts of past data to generate text based on learned predictive distributions. It is a *generative (G)* language model as it stepwise generates *tokens* (think of syllables or words; you can use https://platform.openai.com/tokenizer, (accessed on 15 May 2023) to see how words are split into tokens) to build sentences, paragraphs or articles from the learned predictive distribution of tokens given that were previously generated or provided tokens. The *Chat* in ChatGPT stands for the purpose of the neural network as a conversational chatbot. Its applications can go well beyond that however; see for example Lopez-Lira and Tang (2023). Further general readings on ChatGPT include Ray (2023) or Bahrini et al. (2023).

In this paper, we visit correlation pitfalls in the context of financial and insurance risk modeling and assess whether ChatGPT is capable of handling them. This provides a readable introduction to the topic and clarifies some misunderstandings; for a technical read, see Embrechts et al. (2002).

Several articles addressing ChatGPT's capabilities in different academic contexts recently appeared. For example, Joshi et al. (2023) assesses ChatGPT for undergrad computer science knowledge. Moreover, Wood (2023) addresses ChatGPT's performance on exam questions in accounting; see Flood (2023) for a quick overview of the main findings. In contrast to simply asking for knowledge in the context of homework or an exam problem, we lead a conversation with ChatGPT, thus providing it with the necessary context (a context that a student or practitioner would provide when wanting to learn about correlation). Within that context, we then evaluate ChatGPT's performance. Note that our exact conversation is not reproducible as questions we ask depend on ChatGPT's answers to previous questions and each question's answer is affected by randomness; see also Hofert (2023). However, revisiting some questions by providing increasingly more context allows us to get an impression of how similar conversations would turn out.

Section 2 briefly mentions our technical setup when working with ChatGPT. Section 3 contains our conversation with the chatbot about correlation pitfalls and related concepts. We analyze and rate each of ChatGPT's responses according to a 5-star rating scheme. Section 4 briefly addresses some of the risks one may face when trying to learn technical concepts using ChatGPT. Section 5 provides a summary of our conversation's findings and concluding remarks.

2. The Setup

The conversation in this section took place in May, 2023. Access to ChatGPT (version 3.5) was provided by The University of Hong Kong. ChatGPT's hyperparameters were set as follows (all defaults):

- Max response: 800. The maximal number of tokens used per response (an integer between 1 and 4000).
- Temperature: 0.7. This hyperparameter (a real number between 0 and 1) controls the randomness of the output (from deterministic to creative responses). Low temperature values (0 to 0.3) provide more repetitive and deterministic outputs, medium temperature values (0.3 to 0.7) provide a balance between repetitive and creative outputs and high temperature values (0.7 to 1) provide highly creative and diverse, but potentially less coherent, outputs.
- Top P: 0.95. Similar to the temperature, this hyperparameter (a real number between 0 and 1) affects the randomness of the output, but uses a different method. Low values narrow down the output's token selection to likelier tokens only, and large values allow the output's tokens to range from likely to less likely ones.
- Past messages included: 10. This hyperparameter (an integer between 1 and 20) controls how many past messages are included in each new query, thus providing context for new queries.

Besides a short analysis of each of ChatGPT's responses, we put a 5-star rating in the margins (0 stars = poor answer; 5 stars = excellent answer; half-stars allowed); this is subjective based on our own experience teaching the topics and what we find important that a question's response should address. Furthermore, for readability, we convert ChatGPT's responses to technical questions to LaTeX and format the output slightly to save space. We chose not to ask ChatGPT to provide LaTeX output right away as it is often not required for understanding an answer and we did not want to restrict ChatGPT in its freedom of how to respond. Additionally, note that we did not format our questions for ChatGPT (neither below nor when we asked them).

3. The Conversation and Its Analysis

Hi ChatGPT, I'd like to ask you about correlation in the context of financial and insurance risk modeling.

> Sure, I'd be happy to help! Correlation is an important concept in financial and insurance risk modeling because it measures the degree to which two variables move together. In other words, it measures the strength of the relationship between two variables.
>
> In the context of financial and insurance risk modeling, correlation is particularly important because it helps to quantify the relationship between different types of risk. For example, if two stocks are highly correlated, it means that they tend to move in the same direction at the same time. This can be a problem for investors who are looking to diversify their portfolio because it means that if one stock goes down, the other stock is likely to go down as well.

> Similarly, in the context of insurance risk modeling, correlation is important because it helps to quantify the relationship between different types of risk. For example, if two types of insurance claims are highly correlated, it means that if one type of claim is more likely to occur, the other type of claim is also more likely to occur. This can be a problem for insurance companies because it means that they may be exposed to more risk than they initially anticipated.
>
> Overall, correlation is an important concept in financial and insurance risk modeling because it helps to quantify the relationship between different types of risk. By understanding the degree to which two variables are correlated, investors and insurers can make better decisions about how to manage their risk exposure.

★★★★⯪ Apart from speaking of "variables" instead of "random variables" in the first and last paragraphs, this is an excellent answer.

Can you provide a short, precise definition of the notion of correlation?

> Yes, of course! Correlation is a statistical measure that quantifies the degree to which two variables are related and how they tend to move together. It ranges between -1 and 1, where -1 indicates a perfectly negative correlation, 0 indicates no correlation, and 1 indicates a perfectly positive correlation.

★★★☆☆ This is not a very precise definition. Additionally, the correlation values $-1, 1$ correspond to perfect negative and perfect positive *linear* dependence, respectively, which would be more meaningful to mention than reusing the term "correlation". Moreover, the fact that correlation does not necessarily make sense for all random variables (see later) could have been noted. We can already see from this response that ChatGPT may struggle with mathematical rigor (more later).

What does perfect positive correlation between two random variables mean?

> Perfect positive correlation between two random variables means that they have a strong relationship and tend to move in the same direction. This means that when one variable increases, the other variable also increases, and vice versa. In other words, if the value of one variable goes up, the value of the other variable also goes up, and if the value of one variable goes down, the value of the other variable also goes down.
>
> A perfect positive correlation is represented by a correlation coefficient of +1, which indicates a perfect linear relationship between the two variables. This type of correlation is often seen in situations where two variables are directly related to each other, such as in the case of height and weight, where taller people tend to weigh more than shorter people.
>
> It is important to note that a perfect positive correlation does not necessarily imply causation between the two variables. Correlation only measures the degree to which two variables are related, but it does not indicate whether one variable causes the other.

★★★★☆ We see from the first sentence that ChatGPT struggles to be precise; the rest of the first paragraph also applies to other concepts of dependence, for example concordance. Only in the second paragraph do we see "linear relationship"; the added example is good to have at this point, even if we did not specifically ask for it. The last paragraph stresses that correlation does not necessarily imply causation, another good point to mention here although we did not ask for it. The lack of precision in ChatGPT's answer is made up for by a well-rounded answer overall, stressing other important points.

Can you explain the difference between correlation and concordance?

> Yes, I can explain the difference between correlation and concordance.
>
> Correlation is a statistical measure that quantifies the degree to which two variables are related and how they tend to move together. It measures the strength and direction of the linear relationship between two variables. Correlation can

> range between −1 and +1, where −1 indicates a perfect negative correlation, 0 indicates no correlation, and +1 indicates a perfect positive correlation.
>
> Concordance, on the other hand, is a statistical measure that assesses the degree of agreement between two ordered variables. It measures the level of similarity in the rank ordering of the two variables. Concordance can range between 0 and 1, where 0 indicates no agreement and 1 indicates perfect agreement.
>
> The main difference between correlation and concordance is that correlation measures the strength and direction of the linear relationship between two variables, while concordance measures the level of agreement between two ordered variables. Correlation is used to assess the degree of association between two variables, while concordance is used to assess the degree of agreement between two variables that are ranked in a certain order.
>
> In summary, correlation and concordance are both measures of the relationship between two variables, but they measure different aspects of that relationship. Correlation measures the strength and direction of the linear relationship, while concordance measures the degree of agreement in the rank ordering of the variables.

★★★☆☆ Besides some repetition, we again see the inaccuracy when describing correlation. The notion of rank appearing in the description of concordance is perhaps most intuitive, but not for an explanation of the population version of concordance (we will come back to this later). The range to be limited between 0 and 1 is simply wrong (see below). The fourth paragraph is mostly fine, but the part "[. . .] is used to assess the degree of association between two variables [. . .]" applies to both the notion of correlation and the notion of concordance. The last paragraph is quite better in this regard.

Isn't the range of concordance also the interval from −1 to 1?

> I apologize for the confusion in my previous response. You are correct that the range of concordance is also −1 to 1, just like correlation.
>
> Concordance measures the degree of agreement between two ordered variables, and it can range from −1 to +1. A value of −1 indicates a perfect negative concordance, meaning that the two variables have a perfect inverse relationship. A value of +1 indicates a perfect positive concordance, meaning that the two variables have a perfect direct relationship. A value of 0 indicates no concordance, meaning that there is no relationship between the two variables.

★★⯪☆☆ Except for points already mentioned, the answer is fine overall. However, concluding "no relationship" from a concordance value of 0 is indeed wrong. For example, one can show that Kendall's tau of the uniform distribution on the unit circle is 0 even though there is a strong relationship between the underlying random variables; see, for example, (Embrechts et al. 2002, Proposition 3).

Does 0 correlation mean that there is no relationship betwen the two underlying random variables?

> Yes, that's correct. A correlation coefficient of 0 indicates that there is no linear relationship between the two underlying random variables. In other words, the two variables are not related to each other in a predictable way. This does not necessarily mean that there is no relationship between the variables at all, but rather that any relationship between the variables is not linear.
>
> It's important to note that a correlation coefficient of 0 does not imply that the two variables are independent of each other. Independence means that the occurrence of one event does not affect the occurrence of the other event. A correlation of 0 only indicates that there is no linear relationship between the two variables, but they may still be related in a nonlinear way. Therefore, it's important to consider other statistical measures and techniques to assess the relationship between variables beyond just correlation.

★★★⯪☆ Our question is one of the correlation pitfalls of quantitative risk management. Additionally, according to the first sentence, ChatGPT falls for it. However, in the second sentence, it speaks of "no linear relationship" again, which is correct (yet not what we asked). The third sentence is wrong again, but the fourth finally gets it right. The second paragraph is overall spot on however.

Can you give an example of two dependent random variables with correlation 0?

> Yes, I can give an example of two dependent random variables with correlation 0.
>
> Suppose we have two variables X and Y, where X represents the number of hours a student studies for an exam and Y represents the grade the student receives on the exam. Let's assume that the relationship between X and Y is not linear, but rather follows a U-shaped curve, meaning that students who study too little or too much tend to perform poorly on the exam, while students who study a moderate amount tend to perform well.
>
> If we plot the data for X and Y, we would see a U-shaped curve, which means that there is a relationship between the two variables, but it is not linear. In this case, the correlation coefficient between X and Y would be close to 0, even though the variables are dependent.
>
> So, while the correlation coefficient measures the strength and direction of a linear relationship between two variables, it does not capture any nonlinear relationships that may exist between the variables. In cases where the relationship between two variables is nonlinear, it is important to use other statistical measures and techniques to assess the relationship between the variables.

★★⯪☆☆ ChatGPT again argues with "data" (sample version) instead of random variables (population version) here, as is often done in applied disciplines such as data science or engineering. The question of the existence of certain notions (here: correlation) is then never a problem. With two or more data points, you can always compute a correlation coefficient, even if the data was generated from a joint distribution with margins that have no second moments. An inaccuracy of ChatGPT is the rather vague "U-shaped curve" instead of saying that Y is X^2. Moreover, similarly, a correlation "close to 0" is instead of exactly 0. The last paragraph is fine and somewhat saves this otherwise pedagogically less than ideal response.

What you are describing is the situation where X is, say, N(0, 1) distributed and Y is X^2. Can you derive mathematically that the correlation between X and Y is 0?

> Yes, I can derive mathematically that the correlation between X and Y is 0. The correlation coefficient between X and Y is given by the formula:
>
> $$r = \frac{\operatorname{cov}(X, Y)}{\operatorname{sd}(X) \cdot \operatorname{sd}(Y)}$$
>
> where $\operatorname{cov}(X, Y)$ is the covariance between X and Y, and $\operatorname{sd}(X)$ and $\operatorname{sd}(Y)$ are the standard deviations of X and Y, respectively. Since X is a standard normal variable ($\mathrm{N}(0, 1)$), we know that its mean is 0 and its variance is 1. Therefore, $\operatorname{sd}(X) = 1$.
>
> Now, let's find the covariance between X and Y: $\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ where $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ are the expected values of X and Y, respectively. Since X is standard normal, we know that $\mathbb{E}(X) = 0$. To find $\mathbb{E}(Y)$, we can use the definition of the expected value:
>
> $$\mathbb{E}(Y) = \int y f(y) \mathrm{d}y$$

where $f(y)$ is the probability density function of Y. Since $Y = X^2$, we can substitute Y in terms of X:

$$\mathbb{E}(Y) = \int x^2 f(x) \mathrm{d}x$$

where $f(x)$ is the probability density function of X.

Since X is a standard normal variable, we know that its probability density function is:

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$$

Substituting this into the equation for $E(Y)$, we get:

$$\mathbb{E}(Y) = \int x^2 \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \,\mathrm{d}x$$

This integral can be evaluated using integration by parts, and the result is: $\mathbb{E}(Y) = 1$. Therefore, $\mathbb{E}(Y) = 1$ and $\mathbb{E}(X) = 0$, which means that $\mathrm{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY)$. To find $\mathbb{E}(XY)$, we can substitute Y in terms of X: $\mathbb{E}(XY) = \mathbb{E}(X \cdot X^2) = \mathbb{E}(X^3)$. Since X is a standard normal variable, we know that its third moment is 0. Therefore, $\mathbb{E}(X^3) = 0$.

Substituting these values into the formula for r, we get:

$$r = \frac{\mathrm{cov}(X, Y)}{\mathrm{sd}(X) \cdot \mathrm{sd}(Y)} = \frac{\mathbb{E}(XY)}{1 \cdot \mathrm{sd}(Y)} = \frac{0}{\mathrm{sd}(Y)} = 0$$

Therefore, the correlation between X and Y is 0, even though Y is a function of X. This example shows that correlation only measures the strength and direction of the linear relationship between two variables, and it does not capture any nonlinear relationships that may exist between the variables.

★★★★⯪ Given that ChatGPT's predictive nature has no built-in mathematical logic and thus can fail on mathematical problems by producing "hallucinations" (see Crouse (2023), Rodger (2023) or Hofert (2023)), this answer is fairly impressive. Yet, there are also two flaws here. The first is that ChatGPT's answer uses $\mathbb{E}(Y) = \int y f(y)\,\mathrm{d}y = \int x^2 f(x)\,\mathrm{d}x$, thus the same letter f for the density of Y and that of X. Although it mentions after each formula what f is, using the same letter for both densities is poor from a pedagogical point of view. The second is that the response is fairly long given that all we needed was to realize that $\mathrm{cov}(X, Y) = \mathbb{E}(XY) = \mathbb{E}(X^3) = 0$. The last equation would have been more meaningful to explain, but of course, we can just ask ChatGPT about it.

You simply used that E(X^3) = 0. Why does that hold?

The reason why $\mathbb{E}(X^3) = 0$ for a standard normal variable X is because the standard normal distribution is an odd function. An odd function is a function that satisfies the property $f(-x) = -f(x)$ for all values of x. In other words, if we reflect an odd function about the y-axis, we get the same function with the sign flipped. The standard normal distribution has a probability density function given by:

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$$

If we substitute $-x$ for x in this equation, we get:

$$f(-x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(-x)^2}{2}} = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$$

> Notice that $f(-x)$ is equal to $f(x)$ multiplied by a constant factor (1). This means that the standard normal distribution is an odd function.
>
> Now, let's consider the moment $\mathbb{E}(X^3)$ of the standard normal distribution. Using the definition of the expected value, we have:
>
> $$\mathbb{E}(X^3) = \int x^3 f(x)\,\mathrm{d}x$$
>
> Substituting the probability density function of the standard normal distribution, we get:
>
> $$\mathbb{E}(X^3) = \frac{1}{\sqrt{2\pi}} \cdot \int x^3 e^{-\frac{x^2}{2}}\,\mathrm{d}x$$
>
> This integral can be evaluated using integration by substitution, with $u = x^2/2$ and $\mathrm{d}u = x\,\mathrm{d}x$. The result is:
>
> $$\mathbb{E}(X^3) = \frac{2}{\sqrt{2\pi}} \cdot \int e^{-u}\,\mathrm{d}u$$
>
> Evaluating this integral, we get: $\mathbb{E}(X^3) = -2/\sqrt{2\pi}$. Therefore, $\mathbb{E}(X^3)$ is equal to a constant factor $-2/\sqrt{2\pi}$ multiplied by an odd function, which means that it is equal to 0.
>
> So, the reason why $\mathbb{E}(X^3) = 0$ for a standard normal variable X is because the standard normal distribution is an odd function, and odd functions have odd moments equal to 0.

★⯪☆☆☆ In the first sentence, one should replace "distribution" with "density". The second sentence clarifies that though. The rest of the paragraph is rather lengthy, with the unnecessary "multiplied by a constant factor (1)". Although the integrands are correct (before and after substitution), the result $\mathbb{E}(X^3) = -2/\sqrt{2\pi}$ is wrong (note that $\frac{2}{\sqrt{2\pi}} \cdot \int_0^\infty e^{-u}\,\mathrm{d}u = 2/\sqrt{2\pi}$ but we do not know what integration bounds ChatGPT considered; again, be reminded that ChatGPT is purely predictive and does not have built-in mathematical logic). In the last sentence of the second-last paragraph, ChatGPT argues that the result is 0 because of this (wrong) factor $-2/\sqrt{2\pi}$ (earlier mentioned as $\mathbb{E}(X^3)$ which we wanted to show is 0) being "multiplied by an odd function" (for this to be correct, one would also need to write "multiplied by an integral of an odd function" or the like). The last sentence in ChatGPT's response then just repeats the initial argument (integral of an odd function) without actually showing that this results holds. Overall, we see that once we hit mathematical logic, it is difficult at best to get a meaningful response from ChatGPT.

Does specifying the distribution function F of X, the distribution function G of Y and the correlation coefficient between X and Y uniquely determine the joint distribution function of the random vector (X,Y)?

> Yes, specifying the distribution function F of X, the distribution function G of Y, and the correlation coefficient between X and Y uniquely determine the joint distribution function of the random vector (X, Y). This is known as the Sklar's theorem, which states that any multivariate distribution with continuous marginal distributions can be uniquely represented by its marginal distributions and a copula function that describes the dependence structure between the variables. The copula function is determined by the correlation coefficient between X and Y, which measures the strength and direction of the linear relationship between the variables. Therefore, once F, G, and the correlation coefficient are specified, the joint distribution function of (X, Y) can be obtained by applying Sklar's theorem.

★☆☆☆☆ This is one of the main pitfalls when it comes to correlation and ChatGPT falls for it. The first sentence is, of course, wrong. Then, it becomes subtle. The "This is known as the

Sklar's theorem" is wrong (the notion of correlation does not appear in Sklar's theorem), but the rest of the sentence is fine. Next, "The copula function is determined by the correlation coefficient between X and Y" is wrong. Rather the opposite holds true, as through Hoeffding's lemma (see later), the correlation coefficient depends on the copula of X and Y.

Let us try to see how ChatGPT reacts if we reveal the problem.

But if X is N(0,1) and Y is X^2*, then we just said that X and Y are not independent. However, by Sklar's theorem, I could combine the distribution function F of X and G of Y by the independence copula and would also obtain a joint distribution with margins F and G and correlation coefficient 0. We thus constructed two different joint distributions with the same univariate margins and correlation coefficient 0, which contradicts the uniqueness, doesn't it?*

> You are correct that the example you provided does contradict the uniqueness of Sklar's theorem. This is because Sklar's theorem assumes that the copula function is unique given the marginal distributions and the correlation coefficient between X and Y. However, in the example you provided, the copula function is not unique since there are multiple copula functions that can result in the same correlation coefficient of 0 between X and Y.
>
> In general, the uniqueness of Sklar's theorem relies on the assumption that the copula function is unique, which may not always be the case. This is especially true when the correlation coefficient is close to zero or when the joint distribution has a complex dependence structure that cannot be fully captured by a single copula function. Therefore, while Sklar's theorem is a useful tool for modeling multivariate distributions with continuous marginal distributions, it is important to keep in mind its limitations and to consider alternative methods for modeling complex dependence structures.

★⯪☆☆☆ From the first sentence, we see that ChatGPT runs in the wrong direction entirely. It is not Sklar's theorem that is the problem, but the fact that correlation and margins generally do not uniquely characterize the joint distribution. The second sentence is also wrong in this regard; Sklar's theorem does not *assume* uniqueness, it *states* that it holds, as long as the margins are continuous; more importantly, as mentioned before, the notion of correlation does not appear in Sklar's theorem. The last sentence of the first paragraph is correct though. Note that the second paragraph is also affected by inaccuracies and flaws of a similar nature.

Sklar's theorem doesn't assume uniqueness, it states uniqueness in case the margins are all continuous (which is the case in our example). The problem is that the correlation coefficient itself is not enough to uniquely characterize a joint distribution function (as we saw from our example). Otherwise we would study correlation coefficients as the quantities that uniquely specify the dependence between random variables rather than copulas.

> You are correct that Sklar's theorem does not assume uniqueness, but rather guarantees uniqueness in the case where the marginal distributions are continuous. However, the issue with the example you provided is that the correlation coefficient alone is not sufficient to uniquely determine the copula function, and therefore, not sufficient to uniquely determine the joint distribution function.
>
> As you mentioned, if the correlation coefficient were sufficient to uniquely determine the joint distribution function, we would not need to study copulas as a separate concept. Copulas provide a way to model the dependence structure between random variables beyond the correlation coefficient, which is limited to linear relationships. Copulas allow for modeling more complex dependence structures, such as non-linear relationships or relationships that change depending on the values of the random variables.
>
> Therefore, while Sklar's theorem provides a powerful tool for modeling multivariate distributions with continuous marginal distributions, it is important to

recognize the limitations of the correlation coefficient and to use copulas as a tool for modeling more complex dependence structures.

★★⯪☆☆ Judging from the first paragraph, ChatGPT changed its mind and now suddenly agrees with us; the "However" then does not make much sense. The last paragraph mentioning "limitations" provides useful advice for any mathematical concepts applied in practice.

The correlation coefficient satisfies an invariance principle. Do you know what it says?

Yes, the correlation coefficient satisfies an invariance principle, which states that the correlation between two variables X and Y is unchanged under certain transformations of X and Y. Specifically, the correlation coefficient between X and Y is invariant under linear transformations of X and Y, as well as under translations of X and Y.

More formally, let a and b be constants, and let $X' = aX + b$ and $Y' = cY + d$ be the transformed variables. Then, the correlation coefficient between X' and Y' is the same as the correlation coefficient between X and Y, and can be expressed as:

$$\mathrm{cor}(X', Y') = \mathrm{cor}(aX + b, cY + d) = \mathrm{cor}(X, Y)$$

This invariance principle is useful in practice, as it allows us to compare the strength and direction of linear relationships between variables that have been transformed or shifted. It also helps to simplify calculations and analysis in situations where we need to work with transformed data.

★★⯪☆☆ The first paragraph could have been written in a more compact way (for example, linear transformations already include translations). Together with the second paragraph, ChatGPT makes one mistake though. The results only hold if $a > 0$ and $c > 0$ or if $a < 0$ and $c < 0$; otherwise, the correlation coefficient will change its sign. The last paragraph is fine.

The concordance measures Kendall's tau or Spearman's rho also satisfy an invariance principle. How does it differ from the invariance principle for correlation coefficients?

Kendall's tau and Spearman's rho are rank correlation coefficients that measure the strength and direction of the association between two variables based on the ranks of their values, rather than their actual values. Like the correlation coefficient, Kendall's tau and Spearman's rho also satisfy an invariance principle, but the invariance principle for these measures is slightly different from that of the correlation coefficient.

Specifically, Kendall's tau and Spearman's rho are invariant under any strictly increasing monotonic transformation of the variables. This means that if we apply any strictly increasing monotonic transformation to both variables, the rank correlation coefficient remains the same. For example, if we take the logarithm of both variables, or apply any other strictly increasing monotonic transformation, the rank correlation coefficient remains unchanged.

In contrast, the invariance principle for the correlation coefficient is limited to linear transformations and translations of the variables. This means that the correlation coefficient is only invariant under transformations that preserve the linear relationship between the variables.

Therefore, while the invariance principle for Kendall's tau and Spearman's rho is similar to that of the correlation coefficient, it is more general and includes a wider range of transformations. This makes Kendall's tau and Spearman's rho more robust measures of association than the correlation coefficient, especially when the relationship between the variables is non-linear or when there are outliers in the data.

★★★★⯪ As we pointed out in Hofert (2023), ChatGPT can often produce more meaningful answers when asked about differences between concepts, rather than when asked about the individual concepts separately. When it comes to the invariance principle, ChatGPT indeed does not disappoint. One point (see also below) that it does not get right is the fact that it speaks of "ranks" of the two underlying random variables even though we do not have multiple realizations here (sample version) but rather consider the population version; see also later.

The fact that concordance measures such as Kendall's tau and Spearman's rho satisfy this invariance principle comes from the fact that measures of concordance can be expressed as functionals of the copula of the underlying random variables. And copulas remain invariant under componentwise strictly increasing transformations of the underlying random variables. As functionals of the underlying copula, concordance measures always exist for any pair of continuously distributed random variables. Does the correlation coefficient also always exist for any pair of continuously distributed random variables?

> No, the correlation coefficient does not always exist for any pair of continuously distributed random variables. In general, the correlation coefficient can only be defined for pairs of random variables that have finite second moments, which is not always the case for all continuously distributed random variables.
>
> Specifically, the correlation coefficient between two random variables X and Y is defined as: $\mathrm{cor}(X, Y) = \frac{\mathrm{cov}(X,Y)}{\mathrm{sd}(X)\cdot\mathrm{sd}(Y)} corr(X, Y)$ where $\mathrm{cov}(X, Y)$ is the covariance between X and Y, and $\mathrm{sd}(X)$ and $\mathrm{sd}(Y)$ are the standard deviations of X and Y, respectively. However, if the second moments of X and Y do not exist, then the covariance and the standard deviations are not well-defined, and the correlation coefficient cannot be defined. In such cases, alternative measures of dependence, such as copula-based measures like Kendall's tau or Spearman's rho, may be used to quantify the association between the variables. Therefore, while concordance measures like Kendall's tau and Spearman's rho always exist for any pair of continuously distributed random variables, the correlation coefficient is only well-defined for pairs of random variables that have finite second moments.

★★★★★ This is an excellent answer and finally a correlation pitfall that ChatGPT does not fall for.

Both Spearman's rho and Kendall's tau are actually also correlations, but not of the original random variables (because then they would not necessarily exist as we just said) but of transformations of the random variables. Do you know what this transformation is in the case of Spearman's rho, so which transformation of X from F and Y from G would you need to apply to X and Y so that that the their correlation coefficient becomes Spearman's rho of (X,Y)?

> Yes, you are correct that both Spearman's rho and Kendall's tau are also correlations, but of the ranks of the original random variables, rather than the original random variables themselves.
>
> The transformation used in Spearman's rho involves replacing each observation of X and Y by their respective ranks, which is the position of each observation when the values of X and Y are sorted in ascending order. Specifically, let $R(X)$ and $R(Y)$ denote the ranks of X and Y, respectively, then Spearman's rho can be expressed as:
>
> $$\rho_s = \mathrm{cor}(R(X), R(Y))$$
>
> where cor is the correlation coefficient between $R(X)$ and $R(Y)$.
>
> To calculate the ranks, we first rank the values of X and Y separately, assigning the smallest value rank 1, the second smallest rank 2, and so on. Ties are assigned the average of the ranks that would have been assigned to the tied observations.
>
> Once we have the ranks of X and Y, we can calculate Spearman's rho using the formula above. The resulting value of ρ_s ranges between -1 and 1, with values of -1 and 1 indicating perfect negative and positive rank correlation, respectively, and a value of 0 indicating no rank correlation.
>
> Therefore, Spearman's rho measures the strength and direction of the association between the ranks of X and Y, rather than the original values of X and Y.

★★★★⯪ In principle, this is a good answer. However, it is plagued by the already mentioned focus on "data" (sample versions) rather than the more appropriate theoretical concepts (population versions). Let us see if we can convince ChatGPT to fix that.

I'm not interested in the sample version of Spearman's rho, but the population version, so there is no data of X and Y. Which transformation of X from F and Y from G would you need to apply to X and Y so that that the their correlation becomes Spearman's rho of (X,Y)?

> My apologies for misunderstanding your question earlier. To obtain the population version of Spearman's rho, we need to consider the ranks of the underlying probability distributions F and G from which X and Y are sampled, respectively.
>
> In particular, if we let $U = F(X)$ and $V = G(Y)$ denote the uniformly distributed random variables corresponding to the cumulative distribution functions F and G, respectively, then the ranks of U and V can be calculated as before, by sorting the values of U and V and assigning ranks in the same way as for the sample version. Then, the population version of Spearman's rho can be expressed as:
>
> $$\rho_s = \mathrm{cor}(\mathrm{rank}(U), \mathrm{rank}(V))$$
>
> where $\mathrm{rank}(U)$ and $\mathrm{rank}(V)$ denote the ranks of U and V, respectively, and cor is the correlation coefficient between $\mathrm{rank}(U)$ and $\mathrm{rank}(V)$.
>
> Therefore, to obtain the population version of Spearman's rho of (X, Y), we need to transform X into U using the inverse cumulative distribution function of F, and transform Y into V using the inverse cumulative distribution function of G. We can then calculate the ranks of U and V as before, and compute the correlation coefficient between the resulting ranks to obtain the population version of Spearman's rho.

★★☆☆☆ ChatGPT goes a bit on a roller coaster here. The first paragraph again mentions ranks of X and Y. The second starts as promising by mentioning the correct probability integral transformations $U = F(X)$ and $V = G(Y)$, but then again mentions "ranks" and "values". The third paragraph also contains an error. It is the distribution functions F and G (rather than their inverses) that transform X to U and Y to V, respectively.

The transformations I was looking for were indeed the marginal distribution functions, so the correlation of (U,V)=(F(X), G(Y)) is Spearman's rho. You correctly stated them but then again mention "ranks" and "values". There is no notion of "rank" or "value" if X and Y are random variables. Do you want to rephrase your answer accordingly?

> You are correct, I apologize for the confusion in my previous answer. When X and Y are random variables, we cannot calculate their ranks or values, as these concepts only apply to observed data.
>
> Instead, to obtain the population version of Spearman's rho, we need to transform X and Y into uniformly distributed random variables U and V, respectively, using their marginal distribution functions. Specifically, if F and G are the marginal distribution functions of X and Y, respectively, then we can define $U = F(X)$ and $V = G(Y)$ to obtain two uniformly distributed random variables.
>
> We can then calculate the ranks of U and V by sorting their values and assigning ranks based on their position within the sorted list. Ties are assigned the average of the ranks that would have been assigned to the tied observations. Once we have the ranks of U and V, we can compute the population version of Spearman's rho as:
>
> $$\rho_s = \mathrm{cor}(\mathrm{rank}(U), \mathrm{rank}(V))$$
>
> where $\mathrm{rank}(U)$ and $\mathrm{rank}(V)$ denote the ranks of U and V, respectively, and cor is the correlation coefficient between $\mathrm{rank}(U)$ and $\mathrm{rank}(V)$.

> Therefore, to obtain the population version of Spearman's rho for X and Y, we need to transform X and Y into the uniformly distributed random variables U and V using their marginal distribution functions, and then calculate the ranks of U and V to compute Spearman's rho.

★★⯪☆☆ From the first paragraph, we see that ChatGPT can quickly change its 'mind' (despite being convinced of the opposite in the previous response). The second paragraph is then spot on. However, the third paragraph again mentions the notion of ranks. The last paragraph is a summary sentence, the first part of which is correct, the second again speaking of ranks.

Earlier you mentioned that the correlation coefficient can take on values between −1 and 1. With X being log-normal LN(0,1) distributed and Y being log-normal LN(0,16) distributed, is there a joint distribution for (X,Y) with these margins and correlation 0.5?

> No, there is no joint distribution for (X, Y) with the specified marginal distributions and correlation coefficient of 0.5.
>
> Since X and Y are both log-normally distributed, their joint distribution can be specified using the log-normal bivariate distribution. Specifically, if $(\ln(X), \ln(Y))$ follows a bivariate normal distribution with means (μ_x, μ_y), standard deviations (σ_x, σ_y), and correlation coefficient ρ, then X and Y have a joint log-normal distribution with marginal parameters μ_x and μ_y, and standard deviations σ_x and σ_y, respectively. Using this framework, we can write:
>
> $$\begin{aligned} \ln(X) &\sim \mathrm{N}(0,1) \\ \ln(Y) &\sim \mathrm{N}(0,16) \\ \mathrm{cor}(\ln(X), \ln(Y)) &= 0.5 \end{aligned}$$
>
> To find the joint distribution of (X, Y), we need to transform $\ln(X)$ and $\ln(Y)$ back to their original scale using the exponential function. Letting $Z = \ln(Y)$, we have:
>
> $$\begin{aligned} X &= \exp(\ln(X)) = \exp\big(Z \cdot (\rho\frac{\sigma_x}{\sigma_z}) + \sqrt{1-\rho^2} \cdot \ln(X)\big) \\ Y &= \exp(\ln(Y)) = \exp\big(Z \cdot \sigma_y\big) \end{aligned}$$
>
> where $\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2 - 2 \cdot \rho \cdot \sigma_x \cdot \sigma_y}$ is the standard deviation of Z.
>
> However, since the correlation coefficient between $\ln(X)$ and $\ln(Y)$ is fixed at 0.5, the resulting joint distribution of (X, Y) obtained using this transformation will also have a correlation coefficient of 0.5. Therefore, it is not possible to obtain a joint distribution of (X, Y) with the specified marginal distributions and correlation coefficient of 0.5.

★★☆☆☆ The first sentence is correct and sounds promising. For given margins, Hoeffding's lemma implies that the counter- and comonotone copulas generate the smallest and largest possible correlations and thus, there is an interval of admissible correlations containing 0; see (McNeil et al. 2015, Lemma 7.27). However, the endpoints are not necessarily −1 and 1 as those values would only be attained under perfect linear dependence which is not always possible (it depends on the margins). The second paragraph considers $(\ln(X), \ln(Y))$ to be a bivariate normal distribution with a correlation of $\rho = 0.5$. By the invariance principle, the copula assumed for (X, Y) is, therefore, the normal copula. Thus, the mentioned "log-normal bivariate distribution" for (X, Y) is actually a normal copula with the log-normal margins. However, the correlation matrix with a off-diagonal entry 0.5 that affects the normal copula as a parameter matrix has no meaning anymore as the correlation matrix of (X, Y), since the correlation coefficient of (X, Y) also depends on the marginal distributions. In that sense, the beginning of the third paragraph with "since the correlation coefficient between $\ln(X)$ and $\ln(Y)$ is fixed at 0.5" is wrong. After that point, ChatGPT draws the (correct) conclusion that such a model for (X, Y) does not exist even though it just described (with flawed logic) how to construct such a model. Thus, ChatGPT magically does a U-turn here and comes up with the right conclusion that the desired model for (X, Y) does not exist, even though it describes how (it thinks) the model can be constructed.

Note that directly asking ChatGPT how to construct a model for (X, Y) with $X \sim \mathrm{LN}(0, 1)$, $Y \sim \mathrm{LN}(0, 16)$ and a correlation of 0.5 makes ChatGPT indeed fall for the trap; it then describes a model and claims it has $X \sim \mathrm{LN}(0, 1)$, $Y \sim \mathrm{LN}(0, 16)$ margins and a correlation of 0.5; see Hofert (2023). Additionally here, the logic underlying ChatGPT's response is flawed since ChatGPT assumes the correlation coefficient not to change when we change the marginal distributions.

For X from F and Y from G with finite second moments, there is always an interval of maximal length that contains all possible correlations for (X,Y). Do you know how to find the endpoints of this interval of attainable correlations?

> Yes, the interval of attainable correlations for (X, Y) can be found using the Cauchy-Schwarz inequality. Let X and Y be two random variables with finite second moments, and let μ_x, μ_y, σ_x, and σ_y denote their means and standard deviations, respectively. Then, the correlation coefficient between X and Y is given by:
>
> $$\rho = \frac{\mathrm{cov}(X, Y)}{\sigma_x \cdot \sigma_y}$$
>
> where $\mathrm{cov}(X, Y)$ is the covariance between X and Y. By the Cauchy-Schwarz inequality, we have:
>
> $$|\,\mathrm{cov}(X, Y)| \leq \sigma_x \cdot \sigma_y$$
>
> with equality if and only if there exist constants a and b such that $aX + bY = c$ with probability 1, where c is a constant. Therefore, the correlation coefficient ρ satisfies: $-1 \leq \rho \leq 1$ with equality if and only if there exist constants a and b such that $aX + bY = c$ with probability 1, where c is a constant.
>
> The endpoints of the interval of attainable correlations are achieved when $|\rho| = 1$, which implies that there exist constants a and b such that $aX + bY = c$ with probability 1 for some constant c. In this case, the endpoints are given by:
>
> $$\begin{aligned} \rho = -1 \quad &\text{if } aX + bY = c \text{ for some constants } a \text{ and } b\text{, with } a < 0 \text{ and } b > 0 \\ \rho = 1 \quad &\text{if } aX + bY = c \text{ for some constants } a \text{ and } b\text{, with } a > 0 \text{ and } b > 0 \end{aligned}$$
>
> To find the values of a, b, and c that achieve the endpoints, we can use the method of Lagrange multipliers. Specifically, we can maximize or minimize the function $a^2 + b^2$ subject to the constraint $aX + bY = c$, and use the resulting values of a, b, and c to obtain the endpoints of the interval of attainable correlations.
>
> The details of this calculation will depend on the specific distributions F and G of X and Y, but in general, it can be a challenging problem to solve analytically.

★★☆☆☆ The first paragraph quotes the Cauchy–Schwarz inequality to state that the correlation coefficient lies in $[-1, 1]$. That does not help in finding the largest possible interval of admissible correlations for given margins however. In the beginning of the second paragraph, we see the problem. ChatGPT still believes "the interval of attainable correlations are achieved when $|\rho| = 1$", but then we would always have $[-1, 1]$ as the interval of admissible correlations and thus, perfect linear dependence would always be possible for any margins with finite second moments.

An easier idea is to apply Hoeffding's lemma or identity. Do you know the basic idea?

> Yes, Hoeffding's lemma or identity is another way to find the interval of attainable correlations for (X, Y).

Hoeffding's lemma states that for any two random variables X and Y with finite second moments, we have: $\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ where $\mathbb{E}$ denotes the expected value. Using this lemma, we can rewrite the covariance between X and Y as: $\mathrm{cov}(X,Y) = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. Then, we can use the Cauchy–Schwarz inequality to obtain: $|\mathrm{cov}(X,Y)| \leq \sqrt{\mathrm{var}(X)} \cdot \sqrt{\mathrm{var}(Y)}$ where var denotes the variance. Therefore, we have: $-1 \leq \rho \leq 1$ where ρ is the correlation coefficient between X and Y, and the endpoints of the interval of attainable correlations are given by:

$$\rho = -\sqrt{\mathrm{var}(X)} \cdot \frac{\sqrt{\mathrm{var}(Y)}}{\mathbb{E}[X]-\mathbb{E}[Y]} \quad \text{if } \mathbb{E}[X] \neq \mathbb{E}[Y]$$
$$\rho = [-1,1] \quad \text{if } \mathbb{E}[X] = \mathbb{E}[Y]$$

The first formula gives the endpoints of the interval of attainable correlations when the means of X and Y are different, while the second formula gives the entire interval when the means are equal.

This method is generally easier to apply than the method of Lagrange multipliers, but it assumes that the means and variances of X and Y are known. If they are not known, then we would need to estimate them from a sample, which can introduce additional uncertainty into the calculation.

★☆☆☆☆ The first identity in the second paragraph is not what Hoeffding's lemma says, though it is a straightforward identity and property of the mean. Hoeffding's lemma expresses $\mathrm{cov}(X,Y)$ as a monotone function of the copula C of (X,Y) and thus allows one to conclude that the smallest and largest attainable correlations correspond to the counter- and comonotone copulas, respectively; see (McNeil et al. 2015, Lemma 7.27). The remaining part of the paragraph contains several errors as well.

Clearly, ChatGPT is not suitable for answering this question. We also explained the statement of Hoeffding's lemma to it and asked it to provide us with a stochastic representation of X from $\mathrm{LN}(0,1)$ and Y from $\mathrm{LN}(0,16)$ with a maximal possible correlation for (X,Y). Similarly as before, ChatGPT's answer was lengthy and contained many inaccuracies and errors; hence, we omitted it. ChatGPT's final answer was $(X,Y) = \big(\exp(\Phi^{-1}(U)), \exp(4\Phi^{-1}(V))\big)$ for independent $U,V \sim U(0,1)$. The margins are the correct $\mathrm{LN}(0,1)$ and $\mathrm{LN}(0,16)$ distributions, respectively, but the copula should be the comonotone copula instead of the independence copula.

As a final correlation pitfall, we consider a connection between correlation and the risk measure value-at-risk (VaR). The latter is simply the α-quantile of the underlying distribution (typically for α close to 1).

Does the maximal correlation of (X,Y) imply that the risk measure value-at-risk of the sum X+Y at confidence level alpha in (0,1) is maximal, too?

No, the maximal correlation of (X,Y) does not imply that the risk measure value-at-risk of the sum $X+Y$ at confidence level $\alpha \in (0,1)$ is maximal, too.

The value-at-risk (VaR) is a risk measure that measures the maximum loss that a portfolio of assets can incur at a given confidence level α in $(0,1)$. Specifically, the VaR at confidence level α is defined as the loss that is exceeded with probability α:

$$\mathrm{VaR}(\alpha) = \inf\{x : ¶(X+Y \leq x) \geq 1-\alpha\}$$

where X and Y are the random variables representing the returns of the assets.

The VaR of the sum $X+Y$ depends on the joint distribution of X and Y, not just their correlation. Even if X and Y have the maximal possible correlation, their joint distribution can still have a wide range of shapes and tails, which can affect the VaR of the sum $X+Y$.

For example, consider two random variables X and Y that are perfectly correlated, and let their joint distribution be a bivariate normal distribution with mean zero, standard deviation one, and correlation coefficient one. In this case, the correlation between X and Y is maximal, but the VaR of the sum $X + Y$ at confidence level α is not necessarily maximal. The VaR depends on the tails of the joint distribution, which can be affected by the marginal distributions and other parameters of the bivariate normal distribution.

Therefore, while the maximal correlation between X and Y can have implications for their joint distribution and dependence structure, it does not necessarily imply that the VaR of the sum $X + Y$ is maximal as well. The VaR depends on other factors such as the shape and tails of the joint distribution, and the choice of confidence level α.

★★⯪☆☆ ChatGPT indeed does not fall for this correlation pitfall. By Hoeffding's lemma, the maximal correlation corresponds to the comonotone copula ($M(u,v) = \min\{u,v\}$, $u,v \in [0,1]$) under which one knows, see (McNeil et al. 2015, Proposition 7.20), that VaR is additive irrespective of the margins. However, other copulas for $X \sim F$ and $Y \sim G$ can lead to VaR being superadditive in which case one would obtain a VaR value for $X + Y$ that is even larger than under the comonotone copula. For example, if F and G are $\mathrm{Par}(1/2)$ with the Pareto Type I distribution function $F(x) = 1 - x^{-1/2}$, $x \geq 1$, then the independence copula leads to a larger VaR for $X + Y$ than the comonotone copula; see (Hofert et al. 2020, Exercise 2.28). For an example with all moments finite, take two independent unit exponentials and consider small to moderate confidence levels $\alpha \in (0,1)$. Additionally here, VaR of $X + Y$ under the independence copula is larger than under the comonotone copula.

In the second paragraph, the definition of VaR as "maximum loss that a portfolio of assets can incur at a given confidence level α" is pedagogically not ideal; see also Hofert (2023). The third paragraph also has a subtle flaw: it does not mention that the "wide range of shapes and tails" of the joint distribution of (X,Y) are induced by the margins, because fixing the maximal correlation already means that the underlying copula is the comonotone copula (as mentioned before, by Hoeffding's lemma). The third paragraph indeed describes a model (bivariate normal, correlation 1) that has the comonotone copula, but speaking of "maximal" VaR makes no sense as the margins are already fixed to be $\mathrm{N}(0,1)$. Towards the end of the third paragraph, ChatGPT at least mentions that VaR of the sum $X + Y$ depends (also) on the marginal distributions, a hint in the right direction. Similarly as before, the last paragraph is merely a summary.

4. Risks When Using ChatGPT to Learn about Risk

We asked ChatGPT 22 questions in and around the topic of correlation pitfalls in the context of financial and insurance risk modeling. In doing so, we can identify some of the risks when learning about concepts such as correlation pitfalls from and with a chatbot.

1. *The risk of believing wrong statements.* Sometimes, ChatGPT's answers sound 100% convincing and are still wrong. The problems are not obviously wrong answers, but rather those that can be correct under weak assumptions but are still wrong in general or in the context in which the answer is given. Believing such statements can be disastrous for learning, since, once learned, it will become even harder to identify them as wrong.
2. *The risk of not obtaining meaningful mathematical arguments.* The more technical (mathematical) a question, the more the human interacting with ChatGPT should already know about the topic in order to assess whether the given answer or logical argument is correct (which somewhat defeats the purpose of using ChatGPT for learning new concepts in the first place). Note again here that ChatGPT is purely predictive and does not have built-in mathematical logic.
3. *The risk of establishing trust.* Related to the previous two points, asking more basic, non-technical questions (as one would start a typical conversation with) establishes

some sort of trust in the chatbot to provide one with reliable answers—as one is used to from human conversations. However, one still needs to carefully check every single one of ChatGPT's answers for correctness. Especially tricky is the situation when a given answer is wrong in an obvious manner; one mentions the problem to ChatGPT in the next question, ChatGPT apologizes for being wrong and provides a new answer, and then that new answer is still wrong. The latter, second mistake is often much harder to spot.

4. *The risk of applying ChatGPT in different contexts and expecting the same performance.* Just because ChatGPT turns out to be useful in some applications does not necessarily imply that it is also a meaningful tool in other applications. For example, similar to the aforementioned points, one needs to stay alert concerning ChatGPT's "hallucinations", especially when using ChatGPT in more technical contexts.
5. *The risk of not providing enough context or asking in the right way.* Just because ChatGPT may fail to provide a correct answer does not mean that ChatGPT is not able to provide one. The human interacting with ChatGPT may have simply not provided ChatGPT with enough context or ask a question in a more optimal way for ChatGPT to produce a meaningful response. To do so often requires one to understand the (context of the) question very well, which may somewhat defeat the purpose of asking ChatGPT the said question in the first place (the more interesting questions to ask are those that one does *not* know much about).
6. *The risk of learning sub-optimally.* Even though ChatGPT may provide an entirely correct answer, the underlying argument is sometimes not suitable from a pedagogical point of view (longer or more complicated than necessary, skipping the most important steps, taking an analytical rather than a stochastic approach, etc.). A human teacher typically provides more motivation, focuses on the most important steps, the underlying ideas, etc. Above all, a human teacher demonstrates *passion* about the subject, thus igniting a spark in students to learn more about the topic.

5. Summary and Conclusions

On the one hand, ChatGPT only showed a mediocre performance with an average 5-star rating of 2.84. Understanding pitfalls (or some sort of special or niche cases) is much harder than understanding and answering mainstream questions about correlation. We overall saw that non-technical questions are easier for ChatGPT to get right (those based on more widely known ideas or when concepts are compared) than technical questions (that require mathematical rigor). One general problem when answering technical questions is that ChatGPT's attempts at proofs are mostly of an analytical nature. However, many concepts from the realm of quantitative risk management (such as correlation and its pitfalls) are best understood and often proven stochastically, so via stochastic representations involving random variables. We also identified more general risks when using ChatGPT for learning purposes.

On the other hand, ChatGPT's non-mathematical capabilities are rather amazing to the degree that one often forgets one leads a conversation with a machine or *artificial* intelligence. Oftentimes, ChatGPT would also provide additional interesting facts related to our question that we did not even ask for specifically. Even though we are far away from ChatGPT and other chatbots to replace teachers, ChatGPT can already be used as part of one's learning process in the sense of providing summaries or motivations of well-known and well covered material. Moreover, after you have learned about a topic from more reliable sources, you can start a conversation with ChatGPT and check its output in the same way we did here. It might very well raise an interesting question you have not asked yourself yet and thus act as a study buddy. If future versions are trained on more reliable, mathematical study material, ChatGPT might also provide more accurate responses. It would then be interesting to revisit topics such as correlation pitfalls to check the related improvement. Finally, ChatGPT can also provide some inspiration for learning in general. To see this, we end our article with one more question for ChatGPT and its response.

In business practice, one often sees and hears the term "correlation" even though what is meant is not "linear correlation" but rather "dependence". Write a four line poem about this misnomer that reminds practioners to use each term appropriately.

Correlation is linear,
Dependence is broader still,
Use them with precision,
Or risk a costly bill.

Funding: This research received no external funding.

Data Availability Statement: Data sharing not applicable.

Acknowledgments: I'd like to thank The University of Hong Kong for providing access to ChatGPT while this paper was written.

Conflicts of Interest: The author declares no conflict of interest.

References

Bahrini, Aram, Mohammadsadra Khamoshifar, Hossein Abbasimehr, Robert J. Riggs, Maryam Esmaeili, Rastin Mastali Majdabadkohne, and Morteza Pasehvar. 2023. ChatGPT: Applications, opportunities, and threats. Paper presented at 2023 IEEE Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, USA, April 27–28.

Crouse, Megan. 2023. ChatGPT Cheat Sheet: Complete Guide for 2023. Available online: techrepublic.com/article/chatgpt-cheat-sheet/ (accessed on 9 May 2023).

Embrechts, Paul, Alexander J. McNeil, and Daniel Straumann. 2002. Correlation and dependency in risk management: Properties and pitfalls. In *Risk Management: Value at Risk and Beyond*. Edited by M. Dempster. Cambridge: Cambridge University Press, pp. 176–223. Available online: https://people.math.ethz.ch/~embrecht/ftp/pitfalls.pdf (accessed on 20 September 2022).

Flood, Elizabeth. 2023. ChatGPT Fails Accounting Class. Available online: https://www.cfodive.com/news/chatgpt-fails-accounting-class/648715/?utm_source=Sailthru&utm_medium=email&utm_campaign=Issue:%202023-04-27%20CFO%20Dive%20%5Bissue:49999%5D&utm_term=CFO%20Dive (accessed on 12 June 2023).

Hofert, Marius. 2023. Assessing ChatGPT's Proficiency in Quantitative Risk Management. Available online: https://ssrn.com/abstract=4444104 (accessed on 10 May 2023). [CrossRef]

Hofert, Marius, Rudiger Frey, and Alexander J. McNeil. 2020. The Quantitative Risk Management Exercise Book Solution Manual. Available online: https://github.com/qrmtutorial/qrm/releases/tag/TQRMEB (accessed on 26 March 2023).

Joshi, Ishika, Ritvik Budhiraja, Harshal Dev, Jahnvi Kadia, M. Osama Ataullah, Sayan Mitra, Dhruv Kumar, and Harshal D. Akolekar. 2023. ChatGPT and the future of undergraduate computer science: Challenges, opportunities and recommendations. *arXiv* arXiv:2304.14993v2.

Lopez-Lira, Alejandro, and Yuehua Tang. 2023. Can ChatGPT forecast stock price movements? Return predictability and large language models. *arXiv* arXiv:2304.07619.

McNeil, Alexander J., Rudiger Frey, and Paul Embrechts. 2015. *Quantitative Risk Management: Concepts, Techniques and Tools*, 2nd ed. Princeton: Princeton University Press.

Ray, Partha Pratim. 2023. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems* 3: 121–54. [CrossRef]

Rodger, James. 2023. Google Issues Urgent Warning to Anybody Using ChatGPT. Available online: birminghammail.co.uk/news/midlands-news/google-issues-urgent-warning-anybody-26227871 (accessed on 9 May 2023).

Wood, David A., Muskan P. Achhpilia, Mollie T. Adams, Sanaz Aghazadeh, Kazeem Akinyele, Mfon Akpan, Kristian D. Allee, Abigail. M. Allen, Elizabeth D. Almer, Daniel Ames, and et al. 2023. The ChatGPT Artificial Intelligence Chatbot: How Well Does It Answer Accounting Assessment Questions? *Issues in Accounting Education*. [CrossRef]