Data Analysis for Firm Valuation: Stochastic Modeling of Casino Hold Percentages Using Random Walk and ARIMA*

Karl-Heinz Jöckel[1] and Peter Pflaumer[2]

[1]Medical Faculty, University of Duisburg-Essen, Germany
[2]Department of Statistics, Technical University of Dortmund, Germany

Abstract: This paper presents a stochastic approach to partial firm valuation by analyzing casino hold percentages using Random Walk and ARMA(1,1) models. We establish a mathematical framework that models future cash flows as influenced by current values and random disturbances, enabling us to derive variance formulas for discounted cash flows. By treating casino hold percentages as stochastic variables, we apply the Discounted Cash Flow (DCF) method to evaluate the enterprise value of a casino table under both autocorrelated and non-autocorrelated scenarios. Our findings highlight the critical impact of temporal dependencies on risk assessments, illustrating how ARMA models improve accuracy in variance estimation. To demonstrate the practical application, we analyze monthly hold percentages data from casino games, spanning 2004 to 2024 in Nevada. This research offers valuable insights for financial decision-makers, not only in the gaming industry, emphasizing the importance of considering random fluctuations and advanced modeling techniques in (partial) firm valuation.

Keywords: Partial Enterprise Valuation, ARMA Modeling, Random Walk, Temporal Dependencies, Casino Industry, Risk Assessment, Data Mining, Roulette, Blackjack.

1. Introduction

Stochastic enterprise or firm valuation is a crucial aspect of estimating a company's value, considering uncertainties and risks associated with future cash flows (Casey, 2001; Fenyves and Tarnóczi, 2020). This methodology models future profits as random variables following a probability distribution, enabling a comprehensive analysis of the firm's value amidst unpredictable market conditions. Decision-making under uncertainty has become pivotal in firm valuation, emphasizing the need to consider future results that are both significant and unknown.

Stochastic firm valuation employs statistical distributions to address uncertainties, utilizing the Discounted Cash Flow (DCF) method to sum discounted future cash flows. By treating cash flows as random variables, the distribution of the firm's value is determined, aiding decision-makers in aligning their strategies with their risk attitudes. Neglecting autocorrelation in influencing factors poses a risk of flawed decision-making, highlighting its importance. The valuation process is mathematically represented as:

$$w_T = \sum_{t=1}^{T} v^t g_t$$

* Paper presented at OGIK´2024, 20th Conference on Business Information Science, November, 8-9, 2024, Győr, Hungary.

with w_T = firm value

g_t = free cash flow or free cash flow to firm (FCF, FCFF)

$v = \frac{1}{1+i}$ discounting factor

i = interest rate

$t = 1, 2, ..., T$ period

Jöckel and Pflaumer (2024) introduced a method for estimating the variance of the firm or enterprise value distribution by incorporating temporal dependencies in cash flows using ARMA models. The analysis underscores the importance of considering these dependencies, as neglecting them can significantly increase variance, leading to erroneous decision-making. Utilizing ARMA models enables decision-makers to obtain a more accurate assessment of underlying risks and make informed investment decisions based on a comprehensive understanding of the firm's value distribution. The proposed method provides valuable insights for evaluating the uncertainty associated with future cash flows, enhancing the accuracy of investment decision processes.

In this paper, our focus is on partial stochastic enterprise valuation, specifically the hold percentage - an integral aspect representing the portion of money retained by the casino from player bets. This element significantly influences casino revenues[1]. Our primary objective is to derive the value and distribution of hold percentages for the table game Roulette within the casino. This meticulous analysis equips decision-makers with a valuable tool for understanding the stochastic nature of revenue outcomes at the individual table level. Although our study does not delve into the comprehensive valuation of the entire casino, we underscore the significance of knowing the distribution of hold percentages. This knowledge serves as a crucial means for evaluating the stochastic value of the casino, providing insights into the variability and uncertainties associated with each table's contribution. Decision-makers are thus equipped with a powerful tool to assess and manage the dynamic financial landscape of the casino.

Partial enterprise valuation, in general, refers to the valuation of specific components, segments, or aspects of a business rather than the entire enterprise. It involves assessing the value of individual elements or functions within a company, such as particular business units, assets, projects, or revenue streams. The goal is to gain insights into the contribution of these partial elements to the overall value of the enterprise.

In the context of this paper, the focus on "partial stochastic enterprise valuation" implies a specific examination and valuation of a particular aspect or component of the casino enterprise—in this case, the hold percentages for the table game Roulette. Rather than valuing the entire casino, the analysis concentrates on understanding the stochastic nature of revenue outcomes associated with this specific aspect of the business.

[1] Hold percentage, representing the portion of money retained by the casino from player bets, significantly influences casino revenues. The chief table games offered in Nevada casinos, such as Twenty-One, Craps, Roulette, and Baccarat, are generally favorable to the house, with hold percentages varying over time.
Although we refer to hold percentages, in this paper these values are expressed per 100 currency units (USD) to align with the calculation of present value, which requires scaling based on the total amount wagered.

In the gambling industry, partial firm valuation is especially useful due to the diversity of revenue sources. Here are some examples:

- **Casino vs. Online Gambling**: A company may value its physical casino operations separately from its online gambling platform, as each has different cash flows, risks, and growth prospects.
- **Game-Specific Valuation**: For instance, a firm might value its poker, blackjack, and slot machine operations separately, especially if considering selling or expanding specific game types.
- **Geographic Valuation**: If a gambling company operates in multiple regions (e.g., Las Vegas vs. Macau), each location can be valued separately due to regional differences in regulations, taxes, and customer behavior.
- **Hotel and Non-Gambling Assets**: Many casinos have hotels, restaurants, and entertainment venues, which can be valued independently as they contribute differently to overall revenue and profitability.

Partial enterprise valuation can be valuable for decision-makers seeking to assess the individual contributions, risks, and uncertainties associated with different components of a business. It allows for a more detailed and nuanced understanding of the enterprise's overall value, aiding in strategic decision-making and risk management.

The subsequent sections intensively discuss two models with empirical data: 1) cash flows overlaid by white noise, and 2) cash flows overlaid by an ARMA(1,1) process. These cash flows represent hold percentages in our analysis. We regard hold percentages as random variables and apply the Discounted Cash Flow (DCF) method to analyze their distribution within the casino. This analysis forms the foundation for assessing the distribution of the casino's enterprise value.

The dataset comprises monthly hold percentages of the table game roulette from January 2004 through October 2023 (source: https://gaming.library.unlv.edu/reports/nv_table_hold.pdf). The data from the site was accessed on February 15, 2024. The missing values for April and May 2020, during which casinos were closed due to COVID-19, have been linearly interpolated.

2. Distribution of the Present Value of Hold Percentages

2.1. White Noise Disturbance Model

In this section, we delve into the distribution of the present value of hold percentages within the context of a white noise disturbance model. White noise, in this context, represents a sequence of random and independent disturbances with constant variance. Our exploration of this model aims to unravel how unpredictable and random factors influence the present value distribution of hold percentages.
To accommodate this, the relevant cash flows g at time t: 1, 2, ..., are represented as:

$$g_t = \mu + u_t$$

with $E\left(u_t = 0\right)$

$$E\left(u_t u_s\right) = \begin{cases} \sigma^2 & t = s \\ 0 & otherwise \end{cases}$$

where μ, the mean function, reflects the deterministic part, and u_t represents the stochastic part of the earnings.

In the scenario of an infinite lifespan, the enterprise value is given by:
$w = \lim_{T \to \infty} w_T$ if existing
If the stochastic part of the earnings stream is non-trivial, i.e., $Var(u_t) \neq 0$, the question arises: what is the probability that the random variable w falls within a given interval $[a,b]$. To calculate the probability $P(a \leq w \leq b)$ distributional assumptions about the disturbance variable u_t are required.

The simplest case is that of normal distribution,

$u_t = N(0,\sigma^2)$ for t=1,2,…
This implies that the enterprise value w follows a normal distribution with the expected value:

$$E(w) = \frac{v\mu}{1-v} = \frac{\mu}{i}.$$

and the variance:

$$Var(w) = \frac{v^2}{1-v^2}\sigma^2.$$

The probability that w lies between a and b is given by:

$$P(a \leq w \leq b) = \phi\left(\frac{b-E(w)}{v\sigma}\sqrt{1-v^2}\right) - \phi\left(\frac{a-E(w)}{v\sigma}\sqrt{1-v^2}\right).$$

where ϕ represents the distribution function of the standard normal distribution.

For small interest rates i, the above equation is also approximately valid for arbitrarily independent identically distributed u_t (see Gerber, 1971; Jöckel and Sendler, 1981). Although the assumption of a normal distribution for the residuals should be critically assessed, it is worth noting that this assumption is not as restrictive as it may initially seem.

The normal distribution serves as a suitable mathematical model for many real-world phenomena, assuming that numerous influencing components act independently and additively. As the residuals are indeed influenced by various factors, both internal and external to the company, and in many cases are more or less independent, it is plausible that the normal distribution is an appropriate distributional law in this context. Even if a different distributional law is assumed for the disturbance variables, it does not impact the expected value and variance of the distribution of enterprise value. These parameters remain independent of the residual distributions.

2.2 ARMA(1,1) Disturbance Model

From a practical perspective, the assumption of independent identical distributions, even more so than assuming a normal distribution, can be overly restrictive. In real-world scenarios, economic variables often exhibit temporal dependence, making them dependent random variables. Under the assumption of normal distribution, this implies that the covariance or correlation between cash flows in different periods is nonzero.

To address this limitation and move towards a more realistic enterprise valuation, we introduce a stationary stochastic process for the disturbance variable ut in the form of an Auto Regressive Moving Average (ARMA) (1,1) process:

$$u_t = \phi_1 u_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t \qquad \left[|\phi_1| < 1 \ \ and \ \ |\theta_1| < 1 \right]$$

The first part of the process is referred to as the autoregressive (AR) component, and the second part is the Moving Average (MA) component. Here, ϕ_i and θ_i denote the parameters of the process, and ε_t represents the disturbances assumed to be independent, normally distributed random variables[2].

AutoRegressive (AR) Component ($\phi_1 u_{t-1}$):
The AR component represents the temporal dependence by considering the influence of the past disturbance (u_{t-1}) on the current disturbance (u_t).
ϕ_1 is the parameter that determines the strength and direction of this influence. A higher ϕ_1 value signifies a more significant impact of the past disturbance on the current one, indicating a stronger temporal connection.

Moving Average (MA) Component ($\theta_1 \varepsilon_{t-1}$):
The MA component introduces additional temporal dependencies by incorporating the past error term ε_{t-1} and the current error term (ε_t).
θ_1 is the parameter associated with the past error term, and it governs the influence of the previous error on the current disturbance.
The current disturbance (ε_t) is included directly, reflecting the immediate impact of unpredictable factors on the current state.
Together, these components create a dynamic model that accounts for the influence of both past disturbances and past errors, capturing the temporal dependencies inherent in economic variables. The ARMA(1,1) model strikes a balance by considering both autoregressive and moving average elements, providing a realistic representation of the complex interplay between historical and current factors.

In simpler terms, the ARMA(1,1) model, known for capturing temporal dependencies effectively, acknowledges that the current disturbance u_t depends on the past disturbance u_{t-1} with an autoregressive term, and it is also influenced by the past error term ε_{t-1} with a moving

[2] Details of ARMA and ARIMA modeling can be found, e.g., in Chatfield, 2004; Jöckel and Pflaumer, 2022 used these models to analyze and forecast mortality and excess mortality during the Corona pandemic; in Jöckel and Pflaumer, 1981c, a notable early application of the Box-Jenkins method in German scientific literature, monthly gold prices are forecast.

average term. This widely adopted model is popular for its ability to account for time dependencies in economic variables, providing a robust foundation for realistic enterprise valuation.

From a practical perspective, while using a higher-order AutoRegressive-Moving Average (ARMA) process is one approach to capture extensive temporal dependencies, it comes with a potential drawback. Specifically, employing higher-order processes may necessitate the specification or estimation of a larger number of parameters.

The ARMA(1,1) model strikes a balance, offering an effective representation of temporal dependencies while maintaining simplicity in parameterization. This principle of parsimony ensures a more manageable and interpretable model without sacrificing the ability to capture essential time dynamics in economic variables.
Expected value and variance are: (see Jöckel and Pflaumer (2024) for the derivation of the variance of the enterprise value (also the present value of a stochastic annuity).

$$Ew = \frac{v\mu}{1-v} = \frac{\mu}{i},$$

$$Var(w) = \frac{v^2}{1-v^2}\gamma_0\left(1 + 2\alpha v + 2\alpha\phi_1 v^2 \frac{1}{1-\phi_1 v}\right), \text{ where}$$

$$\gamma_0 = Var(u_t) = \frac{1 + 2\phi_1 \cdot \theta_1 + \theta_1^2}{1-\phi_1^2} \cdot \sigma_\varepsilon^2$$

and $\alpha = \dfrac{(1+\phi_1\theta_1)(\phi_1+\theta_1)}{1+\theta_1^2+2\phi_1\theta_1}$.

The variance of the enterprise value is influenced not only by the variance of the disturbance process but also by its covariances. Neglecting temporal dependence in performance indicators can result in a significant underestimation of the enterprise value's variance, especially in the presence of positive autocorrelation. In economic variables, positive autocorrelation is more likely than negative autocorrelation. The extent of this underestimation is illustrated in Jöckel and Pflaumer (2024).

The statement "positive autocorrelation is more likely than negative autocorrelation" refers to a general observation in economic time series data, where consecutive observations are more likely to be positively correlated rather than negatively correlated.

Explanation:

Trends and Persistence: Economic variables often exhibit trends and persistency over time. For example, if a particular economic indicator shows growth in one period, there is a higher likelihood that it will continue to show growth in the subsequent period. This positive correlation reflects the persistence of trends.

Economic Factors: Economic systems are influenced by various factors that tend to have long-term effects. For instance, positive economic conditions, technological advancements, or policy changes may lead to sustained positive trends in economic variables.

Reference to Literature: This observation is grounded in empirical studies of economic time series data. Economic researchers and statisticians have found that positive autocorrelation is a common characteristic in various economic indicators.

While it's a general trend, it's essential to note that this observation may not hold true for all economic variables or in all economic scenarios. It's a tendency rather than an absolute rule. Economic data can be complex, and various factors can influence the correlation structure differently.

We need the parameters of the ARIMA(1,0,1) model to estimate the variance of the firm value. There are two possibilities for obtaining these parameters:

1. **Statistical estimation**, which requires long time series data that are not always available, or
2. **Subjective assignment**, as proposed in Jöckel and Pflaumeer (2024). This article includes a table of different combinations of the two parameters, illustrating the resulting magnitude of the variance.

Here are some possible ways for assigning parameters in the context of the ARIMA(1,0,1) model:

1. **Expert Judgment**: Relying on the insights and experience of subject matter experts to suggest reasonable values for the parameters based on their knowledge of the industry or specific context.
2. **Historical Data Analysis**: Using available historical data to identify patterns or relationships that can inform the selection of parameter values, even if the data set is not long enough for robust statistical estimation.
3. **Literature Review**: Consulting existing studies or articles to find commonly used parameter values in similar contexts or industries, which can serve as a basis for assignment.
4. **Simulation Studies**: Conducting simulation studies to explore how different parameter values impact the model's outcomes, helping to identify values that lead to desired characteristics in the model's behavior.
5. **Sensitivity Analysis**: Assessing how changes in parameter values affect the model's results, which can provide guidance on which parameter values are most critical and help identify suitable assignments.
6. **Use of Default Values**: Adopting commonly accepted default values for parameters that have been established in previous research or practice, especially in the absence of specific data.

3. Case Study on Roulette Casino Hold Percentages

3.1. Data

For our case study, we utilized monthly hold percentages data from the table game roulette spanning the period from January 2004 to October 2023 in Nevada. The data were sourced from https://gaming.library.unlv.edu/reports/nv_table_hold.pdf .

Hold percentages, representing the portion of money retained by the casino from player bets, are critical indicators of a casino's financial performance. The chosen dataset offers a comprehensive view of how these hold percentages fluctuate over an extended timeframe, allowing for a robust analysis of their stochastic nature.

The analysis in this paper has been conducted using the R software.

3.2. Methodology

In conducting our case study, we applied a rigorous methodology to derive valuable insights into the stochastic behavior of casino hold percentages. The following steps outline our approach:

1. Collection and Preprocessing:

Acquired monthly hold percentages data for the table game roulette.
Ensured data integrity and completeness through thorough preprocessing, addressing any missing or erroneous values.

2. Stochastic Modeling:

Modeled the hold percentages as random variables, considering the white noise disturbance model and the ARMA(1,1) process.
Utilized the Discounted Cash Flow (DCF) method to determine the distribution of the casino's enterprise value based on the stochastic nature of hold percentages.

3. Parameter Estimation:

Estimated the parameters of the chosen models, ensuring their alignment with the characteristics of the observed data.
Applied statistical techniques to enhance the accuracy of parameter estimation, considering the temporal dependencies inherent in casino cash flows.

4. Variance Analysis:

Investigated the impact of autocorrelation on the variance estimation of the enterprise value.
Compared scenarios with non-autocorrelated and autocorrelated hold percentages to highlight the importance of realistic modeling in risk assessment.

5. Results and Discussion:

Presented the expected values and variances under different ARMA models, emphasizing the implications for decision-makers in the casino industry.
Discussed practical considerations for parameter estimation and the potential consequences of assuming independence in stochastic enterprise valuation.

Through this methodology, our case study aims to provide valuable insights for decision-makers in the casino industry, offering a nuanced understanding of the stochastic aspects of hold percentages and their role in shaping the financial landscape of casinos. The findings contribute to informed investment decisions amid uncertainty, facilitating a more accurate assessment of the risks associated with future cash flows.

3.3 Time Series Analysis of the Hold Percentages

a) Descriptive Data Analysis

The hold percentages in Figure 1 exhibit fluctuations around a mean of 18.93, with a standard deviation of 3.15. Ranging from a minimum of 9.17 to a maximum of 25.70, the data highlights variability around the central tendency, suggesting a degree of volatility in observed values. Notably, the presence of seasonal components contributes to these fluctuations, indicating systematic variations at regular intervals. This observation sets the stage for a comprehensive time series analysis. The visual representation helps in observing trends, seasonality, and any apparent patterns in the data.

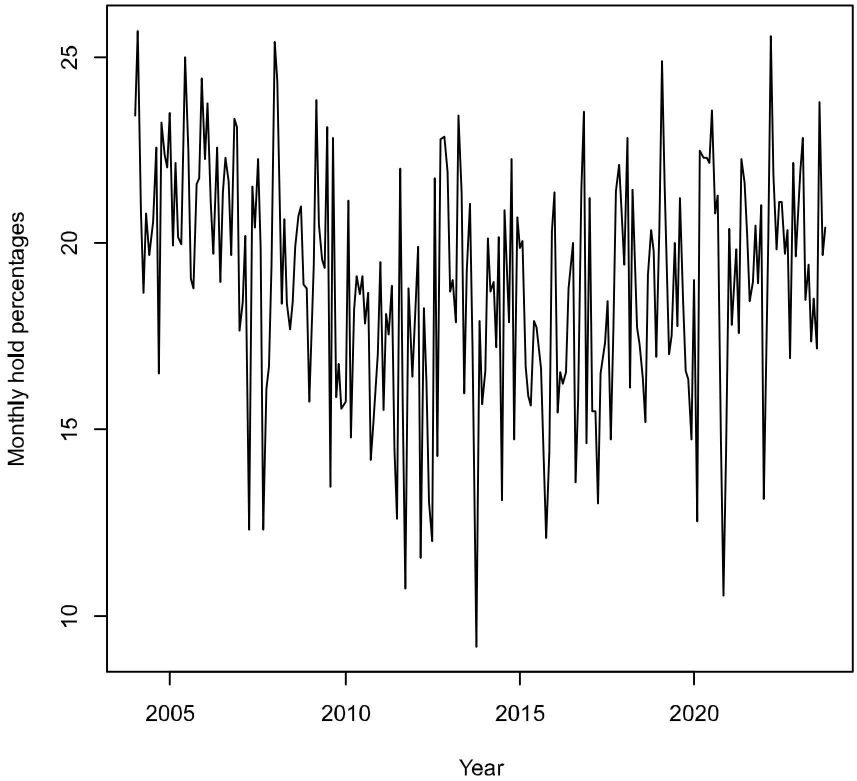


Fig. 1: Monthly hold percentages of Roulette in Nevada from January 2004 to October 2023

Table 1 and Table 2 present monthly statistics of the hold percentage.
These means represent the average monthly hold percentages for the specified periods, offering an overview of the central tendency of the data. The standard deviation indicates the variability or dispersion around the mean.

Table 1: Monthly statistics of hold percentages between 2018 and 2023

Month	N	Mean	SD
1	6	19.06	3.08
2	6	20.16	4.48
3	6	20.68	3.29
4	5	19.46	2.19
5	5	18.94	2.11
6	6	20.11	1.89
7	6	19.37	2.74
8	6	19.85	2.90
9	6	19.63	1.04
10	6	18.24	2.49
11	5	17.55	4.42
12	5	17.34	2.95
Overall	68	19.25	2.87

Table 2: Monthly statistics of hold percentages between 2004 and 2023

Month	N	Mean	SD
1	20	19.678	2.9965
2	20	19.961	3.4966
3	20	19.016	3.3443
4	19	18.465	2.8146
5	19	18.888	2.1233
6	20	18.906	2.669
7	20	18.833	3.3543
8	20	19.238	3.0764
9	20	18.101	2.7037
10	20	18.043	4.2053
11	19	18.789	3.6271
12	19	18.816	3.1011
	236	18.897	3.1431

Missing values (Covid-Lockdown) April 2020 and May 2020.

There appears to be a seasonal pattern in the mean of the monthly hold percentages, with above-average values at the beginning of the year and below-average values towards the end of the year. This pattern is more pronounced in Table 1 (2018-2023) than in Table 2 (2004-2023).

The ascending trend in the means from the beginning to the middle of the year in both tables suggests a potential seasonal effect. However, the decrease in mean values towards the end of the year, especially in Table 1, further supports the observation of a seasonal pattern. To confirm this pattern statistically, we conducted a time series decomposition.

It's important to note that the monthly variations in average hold percentages cannot be entirely explained using methods of time series analysis, as the hold percentage of a table game depends on factors such as average bet, rounds per hour, time played, and house edge.

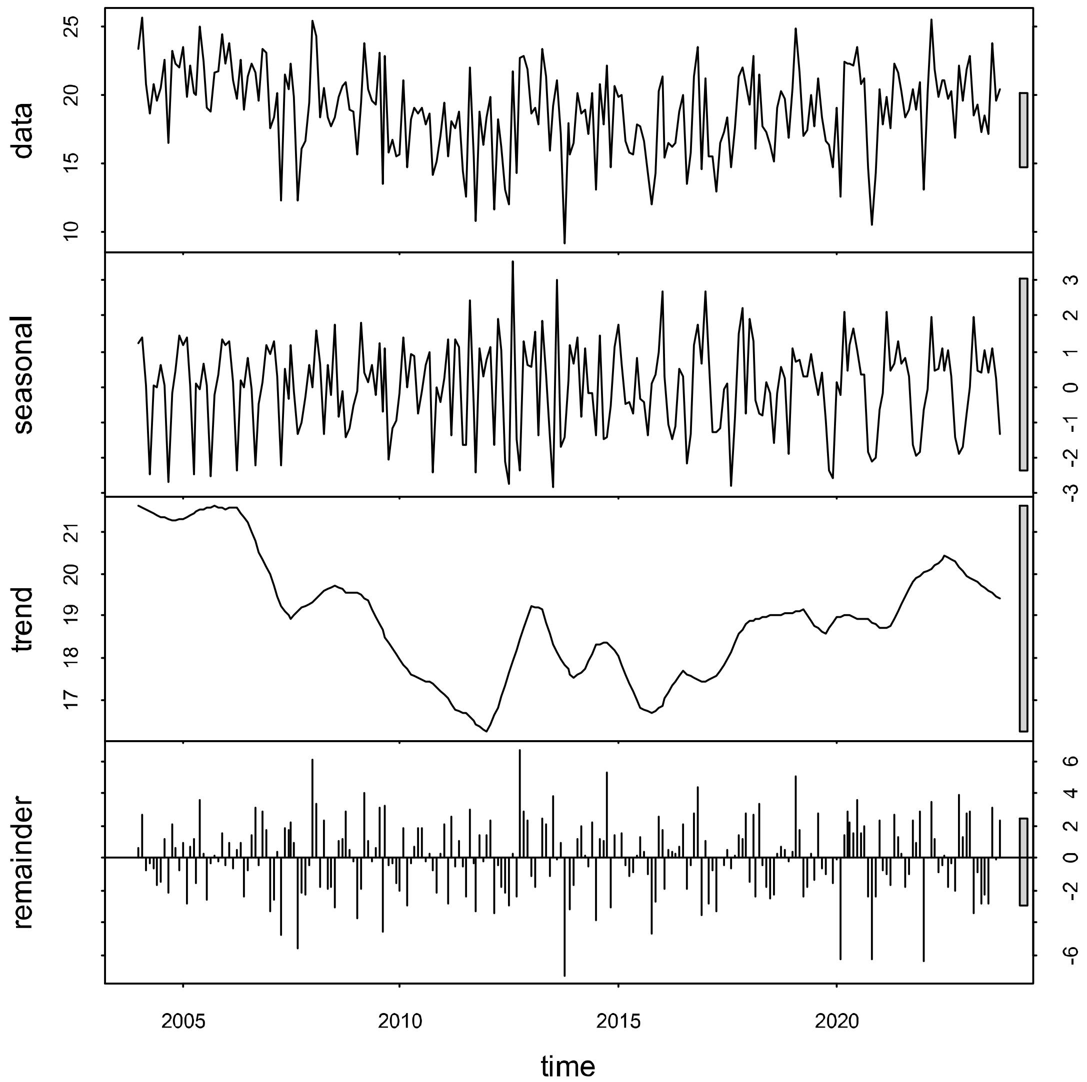


Fig. 2: Time Series Decomposition into Trend, Season, and Remainder for Monthly Hold Percentages (2004-2023)

In Fig. 2, the decomposition of the time series into trend, season, and remainder is illustrated. The trend shows a general decrease from 2004 to 2013, followed by an increase in cycles. The seasonal pattern of the time series is more pronounced in the first and last few years than in the middle of the total period.

b) ACF and PACF

Autocorrelation Function (ACF): In ARMA models, the Autocorrelation Function (ACF) assesses the correlation between a time series and its own lagged values, capturing both direct and indirect dependencies. Peaks in the ACF plot at specific lags indicate potential autoregressive components within the data.
Partial Autocorrelation Function (PACF): The Partial Autocorrelation Function (PACF) is instrumental in ARMA models for discerning the direct relationship between a specific lag and the present observation, while mitigating the influence of intermediate lags. It aids in

determining the order of the AutoRegressive (AR) process, offering insights into the number of lagged terms to incorporate into the model.

To identify potential autocorrelation and partial autocorrelation patterns, we present Fig. 3 depicting the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the mean-adjusted series y=Y-mean(Y). The mean-adjusted series provides insight into the pattern of the disturbances. These plots serve as a guide for selecting ARMA model parameters.

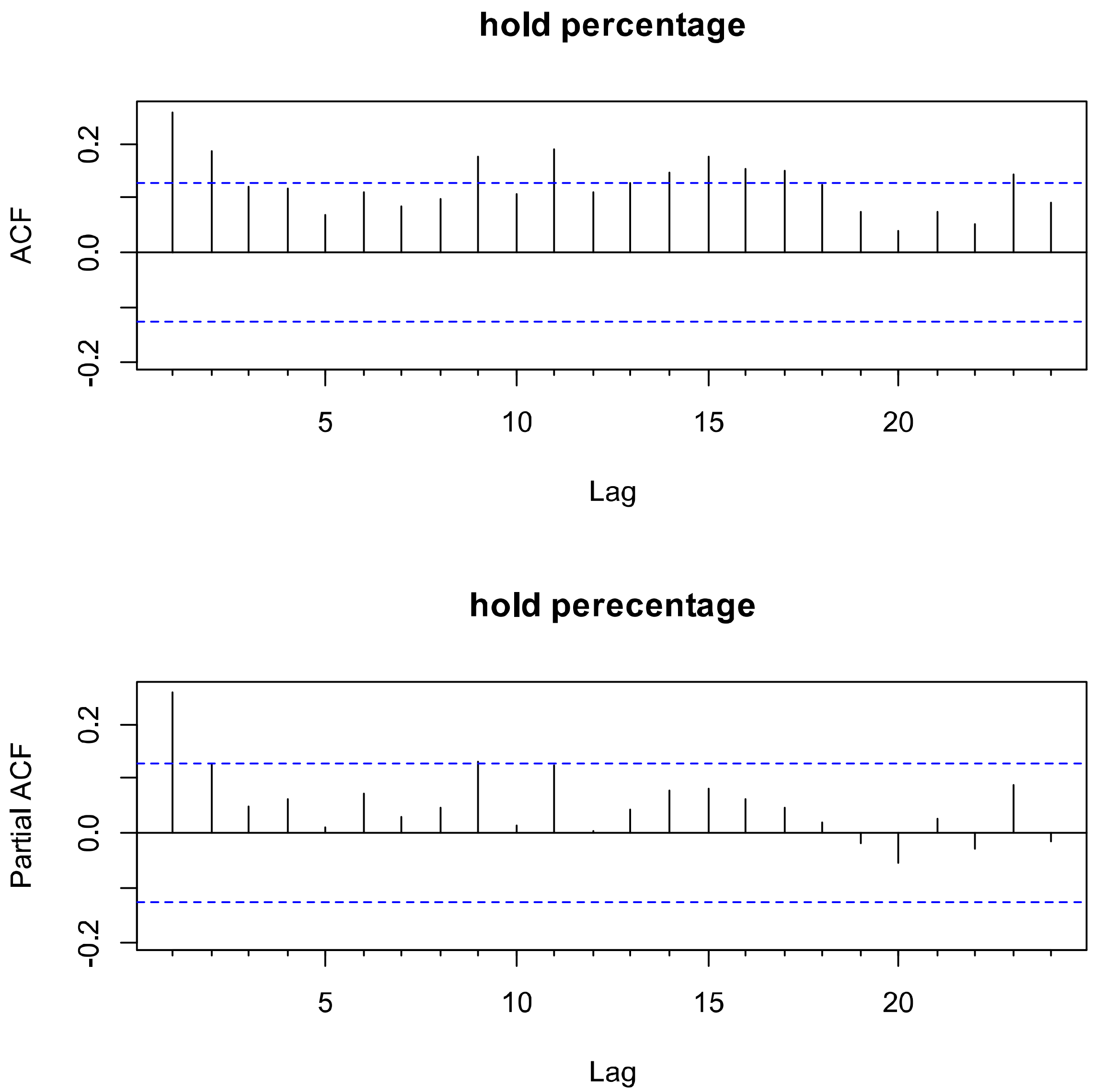


Fig. 3: Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

Examination of the ACF plot, similar to the original series plot, indicates that no differencing of the series is necessary. This implies that we can utilize the original series of hold percentages. The combined analysis of ACF and PACF does not rule out the possibility of an ARMA(1,1) process for hold percentages. In an ARMA(1,1) model, the theoretical ACF exhibits a gradual decay with a notable drop at lag 1, while the PACF displays a distinct spike at lag 1 followed by a gradual decline. This observed pattern suggests a potential level of seasonality in the series, which we will further explore by employing a seasonal ARMA model.

c) Estimation of Parameters, Variance, and AIC of the Mean Adjusted Series y

Upon determining the order of the model, the next step involves estimating the parameters. The statistical software R utilizes maximum likelihood estimation (MLE), a method that identifies parameter values maximizing the likelihood of obtaining the observed data. R provides the log probability of the data, representing the logarithm of the probability that the observed data stems from the estimated model.
The estimation of our model yields the following parameter estimates:

```
arima(x = y, order = c(1, 0, 1))

Coefficients:
         ar1      ma1  intercept
      0.9813  -0.9066     0.4626
s.e.  0.0196   0.0408     0.9093

sigma^2 estimated as 8.78:  log likelihood = -596.54,  aic = 
1201.07
```

The estimates of the AR(1) and MA(1) parameters are found to be significant; however, the intercept is not statistically significant.
While various ARMA and ARIMA models were considered, specified, and estimated, the principle of parsimony was consistently observed. In a set of predictive models, the principle of parsimony (Ockham′s Razor[3]) dictates choosing the simplest models. Three additional models are presented here:

A) `ARIMA(2,1,1)  (Using auto.arima)`[4] `for series y`

```
Coefficients:
         ar1     ar2      ma1
      0.1241  0.0357  -0.9389
s.e.  0.0683  0.0679   0.0218

sigma^2 estimated as 8.83:  log likelihood=-593.81, AIC=1195.62
```

B) `arima(x = y, order = c(1, 1, 1))`

```
Coefficients:
         ar1      ma1
      0.1247  -0.9357
s.e.  0.0685   0.0217

sigma^2 estimated as 8.728:  log likelihood = -593.95,
  aic = 1193.9
```

C) `arima(x = y, order = c(1, 0, 1), seasonal=list(order = c(1, 0, 1),` **`period = 12`**`), method = "ML")`

```
Coefficients:
         ar1      ma1   sar1     sma1  intercept
      0.9837  -0.9056  0.834  -0.8904     0.3761
s.e.  0.0175   0.0379  0.191   0.1692     0.7708
sigma^2 estimated as 8.694:  log likelihood = -595.58,
  aic = 1203.15
```

[3] Ockham's Razor: Named after the scholastic philosopher William of Ockham (c. 1287–1347);
this principle suggests that the simplest explanation or model with the fewest assumptions is preferable.
[4] auto.arima is a function in the R programming language, specifically within the forecast package. This function is used for automatically selecting the best ARIMA model for a given time series.

We also considered an ARIMA(1,0,1)(1,0,1)[12] or ARMA(1,1)(1,1)[12] model for the disturbances, denoted as u(t). Specifically, ARIMA(1,0,1)(1,0,1)[12] denotes a model with an autoregressive component of order 1, no differencing (d=0), a moving average component of order 1, a seasonal autoregressive component of order 1, no seasonal differencing (D=0), a seasonal moving average component of order 1, and a seasonal period of 12 (assuming monthly data).

The estimation results excluding the intercept are significant. When comparing with the non-seasonal ARMA(1,1) model, the variance is slightly smaller, but the AIC has slightly risen, indicating that the more complex model specification does not significantly improve the model quality.

However, our primary focus is on the valuation of the hold percentages, and for this purpose, we adhere to the ARMA(1,1) model. The consideration of other models becomes pertinent when the primary objective is the forecasting of future hold percentages.

Akaike's Information Criterion (AIC) is employed to determine the order of an ARIMA model. AIC facilitates the comparison of different models, selecting the one that best fits the data. The optimal model, according to AIC, explains the greatest variation using the fewest independent variables. Minimizing AIC results in good models (see, for example, Hyndman and Athanasopoulos, 2018). An earlier precursor to AIC is Theil's adjusted coefficient of determination.

For comparison, the estimation results of the white noise model are presented. White noise implies uncorrelated disturbances, and the observations swing randomly around the mean of 18.93, indicating a lack of systematic patterns or dependencies in the data. However, the AIC of the white noise model is greater than the AIC of the ARMA(1,1) model. This suggests that considering dependencies in the ARMA(1,1) model improves the model's quality compared to the simpler assumption of uncorrelated disturbances.

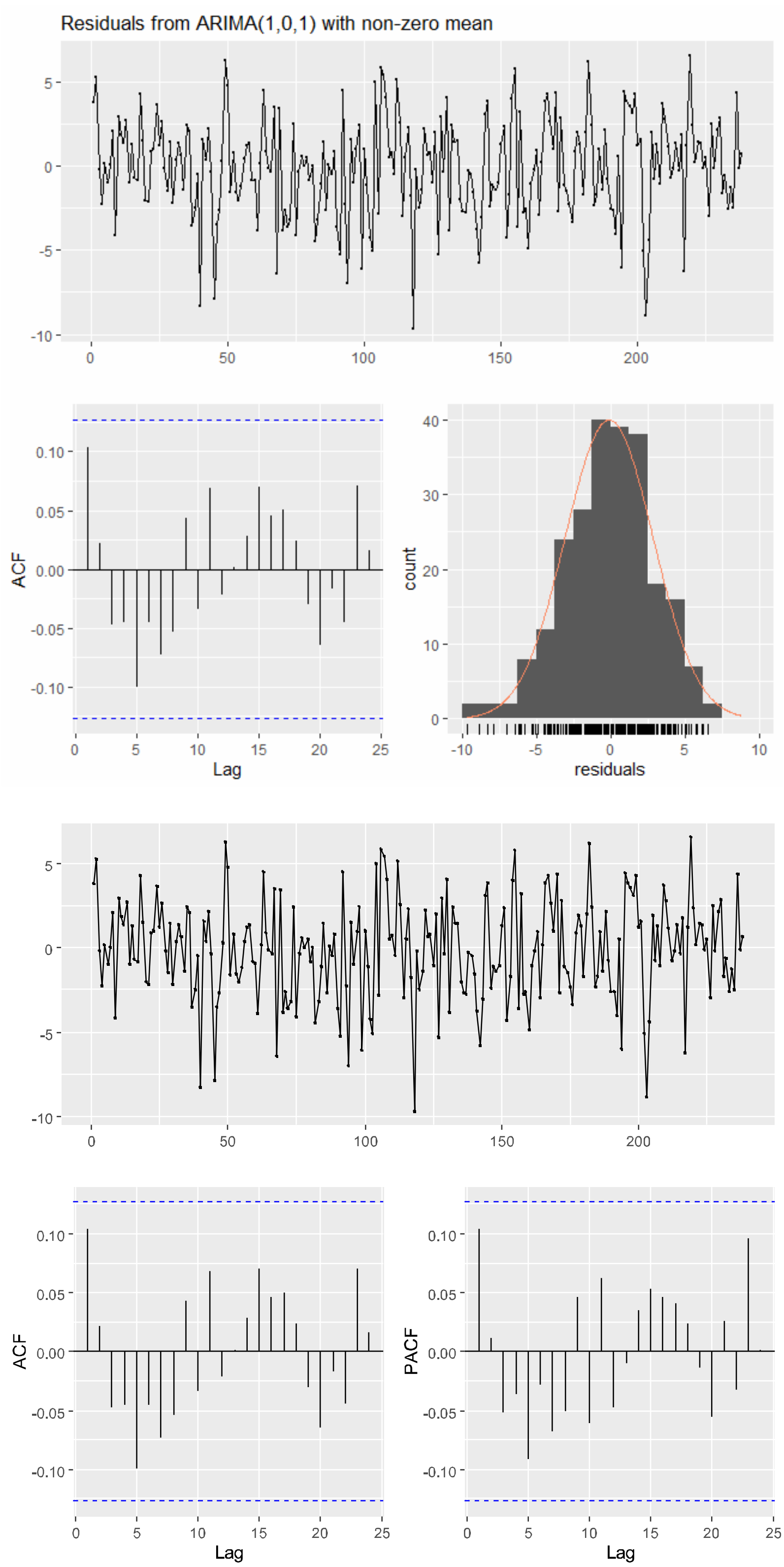
White noise of the Original Values Y:

```
Coefficients:
      intercept
        18.9259
s.e.     0.2035

sigma^2 estimated as 9.852:  log likelihood = -609.94,
  aic = 1223.87
```

d) Analysis of Residuals

We examine the residuals to ensure they exhibit no discernible patterns or trends. The model residuals, shown in Fig. 4a and 4b, display no significant spikes; however, a potential seasonal pattern is noted. Additionally, the residuals demonstrate a normal distribution, contributing to the adequacy of our model specification. All in all, we believe that our model is adequately specified.



Figs. 4a and b: Graphical results of the residual analysis

e) Forecast Plots

We generate forecast plots based on the ARMA model to visualize the expected trajectory of hold percentages into the future (see Figure 5). The point forecasts converge quickly, tending towards the mean of the stochastic process, which is 0 considering the mean-adjusted series. The forecast intervals capture the variations observed in the past.

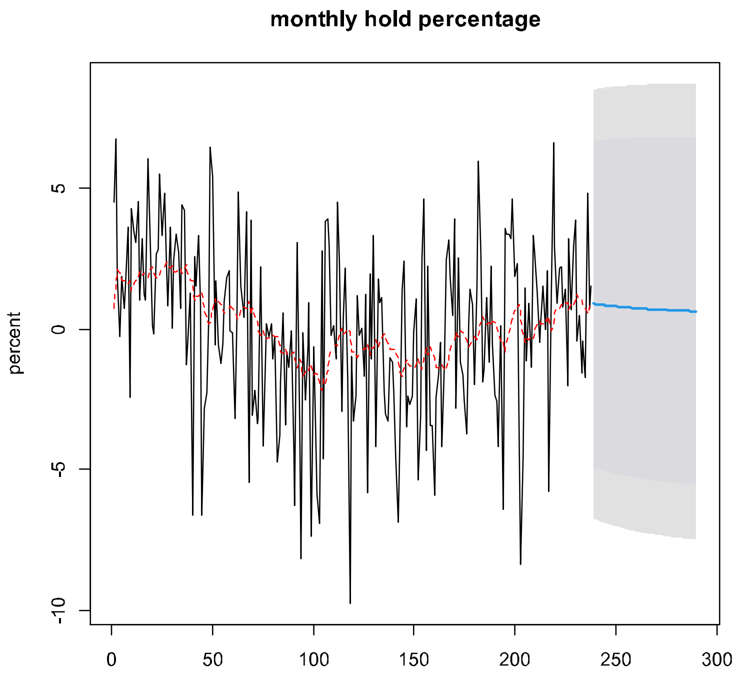


Fig. 5: Mean adjusted hold percentages (actual and fitted) and point forecasts with 95%- and 99%-prediction intervals (black: actual; red: fitted; blue: forecast)

f) Present Value Distribution

Assuming a monthly interest rate of 1%, we calculated the present value distribution of the hold percentages using the estimated ARMA parameters. The expectation E(w)=1892.59 (average hold/i) and the variance is Var(w)=7275.3 (standard deviation = 85.3). For comparison, the white noise model exhibits the same expectation and a standard deviation of 22.13.

Although we discuss hold percentages, it is important to note that these percentages are expressed per 100 currency units (USD). To calculate the present value distribution, both the mean and standard deviation must be multiplied by the actual total amount wagered, divided by 100. For example, if the average total monthly bets are 10,000 USD, the mean and standard deviation of the hold percentages should be multiplied by 100 (i.e., 10,000/100) to obtain the final present value distribution. Additionally, the risk considered in this analysis focuses solely on the variability of the hold percentages and does not account for fluctuations in the total amount wagered each month. For the purposes of this model, we assume the total bets remain constant, and our focus is on how variations in hold ratios influence the present value distribution.

In Figure 6, the normal density functions of the present value of the hold percentages are illustrated.

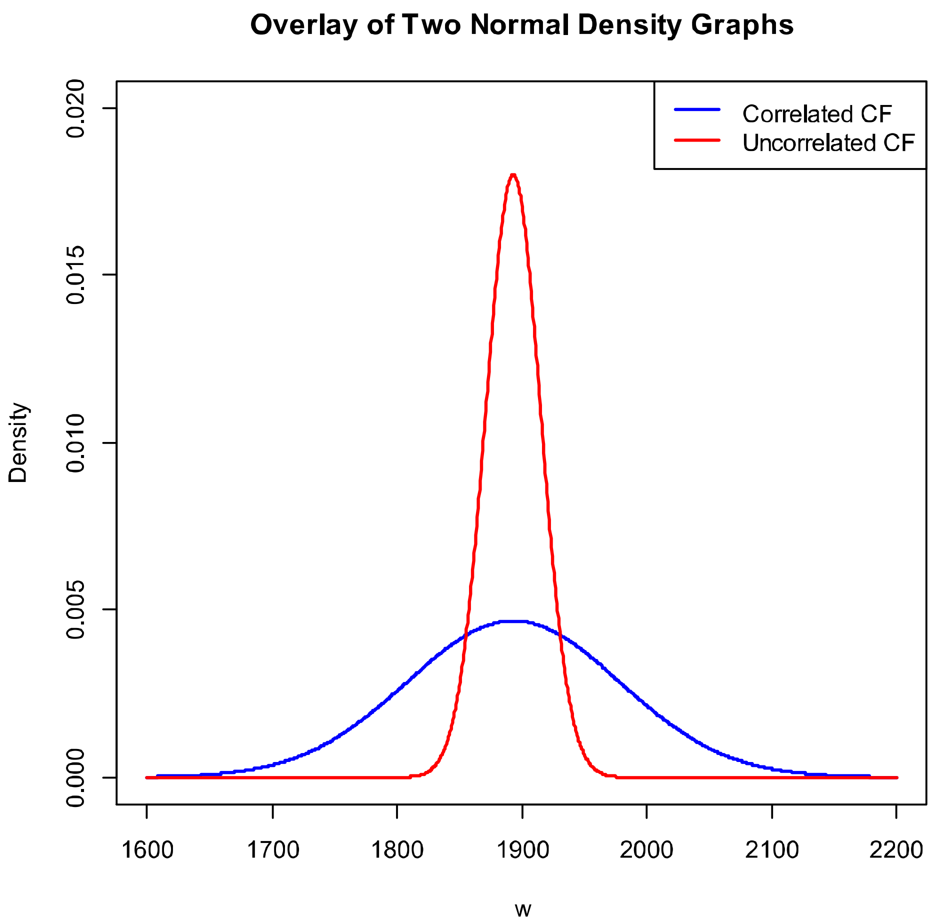


Fig. 6: Distribution of the present values of the hold percentages
blue: N(1893; 85.3) and red: N(1893,22.1)

The risk-averse investor may make a mistake if they fail to consider the dependencies inherent in the hold percentages. In the context of the ARIMA(1,1) model versus the white noise model:

ARIMA(1,1) Model (N(1893, 85.3)):
Captures temporal dependencies and autocorrelation in the data.
Higher standard deviation (85.3) indicates a more complex and variable stochastic process.
Reflects a more realistic representation of the potential future values of hold percentages, considering historical patterns.

White Noise Model (N(1893, 22.1)):
Assumes independence and lack of temporal dependencies in the disturbances.
Lower standard deviation (22.13) implies less variability and a simpler, uncorrelated structure.
May underestimate the true variability and risk associated with future hold percentages.

For a risk-averse investor, neglecting dependencies might result in an inaccurate assessment of the potential risks and uncertainties in the casino industry. The ARIMA(1,1) model, by incorporating dependencies, provides a more nuanced and likely more accurate representation of the stochastic behavior of hold percentages. Therefore, the investor's decision-making process should involve a careful consideration of the temporal dependencies captured by the ARIMA(1,1) model to make informed and prudent investment decisions.

4. A Quick and Simple Approach to Assessing Firm Valuation

While ARIMA(1,0,1) is a complex procedure that requires advanced statistical knowledge and a fairly long time series for accurate identification and parameter estimation, it can be challenging for practical application. In contrast, assuming a random walk for cash flows offers a much simpler alternative. This approach allows us to use straightforward formulas that depend primarily on the interest rate and the number of time periods to construct the present value distribution and assess risk. This simplicity makes it a more practical method for firm valuation, especially when time series data is limited or when a quick assessment is needed.

Random Walk Overview

A random walk describes a stochastic process in which each step depends on the previous one, plus a random shock. In financial terms, it models the cash flows where each period's value is the sum of the previous value and a random fluctuation. Mathematically, it can be expressed as:

$$u_t = u_{t-1} + \varepsilon_t$$

where u_t the cash flow at time t, and ϵ_t is an independent and identically distributed (i.i.d.) random shock with mean $E[\epsilon_t]=0$ and variance σ^2. This process reflects the unpredictability of financial projections, as each cash flow fluctuates randomly over time without a clear trend.

Key Formulas for Present Value and Risk Assessment

For firm valuation, the variability in cash flows modeled as a random walk affects the present value distribution. We derived three important formulas that provide insights into the variance of discounted cash flows (see Jöckel and Pflaumer, 2025). These formulas account for the stochastic nature of cash flows and their impact on firm valuation:

1. **Variance of Discounted Cash Flows:**

   The variance of the present value when cash flows are discounted over n periods at a rate q=1+i (where i is the interest rate) is given by:

$$Var(discounted) = \frac{n + \frac{q^{2n+2} - q^2}{q^2 - 1} - \frac{2 \cdot q\left(q^n - 1\right)}{q - 1}}{q^{2n} \cdot (q-1)^2} \cdot \sigma^2$$

   This formula captures the increasing uncertainty over time, where the variability of future cash flows grows as the number of periods increases.

2. **Variance of Undiscounted Cash Flows:**

   When the discount rate approaches zero (i.e., cash flows are not discounted), the variance of total cash flows over n periods is:

$$Var(undiscounted) = \frac{n(n+1)(2n+1)}{6}\sigma^2$$

   This highlights how, without discounting, the cumulative uncertainty grows rapidly as the time horizon extends.

3. **Variance for an Infinite Time Horizon:**

   For an infinite time horizon, the variance of the present value of discounted cash flows approaches:

$$Var(infinite) = \frac{q^2}{(q+1)(q-1)^3}\sigma^2$$

   This expression simplifies the assessment of risk over an extended period and shows how discounting reduces the variance, but the inherent randomness of cash flows still leads to increasing uncertainty.

Example: Casino Hold Percentages

To illustrate the practical application of these formulas, we analyze monthly hold percentages data from the table game Twenty One (Blackjack) in Nevada, spanning from January 2004 to June 2024. The hold percentages represent the portion of money retained by the casino from player bets, a key indicator of financial performance. Using a random walk model for these cash flows, we assume the actual mean hold percentage of 12.73% and apply a monthly interest rate of 2% for discounting. The random walk error term σ^2 is estimated as 2.545 based on the data.

The variance of the present value distribution is calculated for different scenarios:

- **Discounted Case (n = 240, q = 1.02):**
  - Variance = $62{,}239.67 \cdot 2.545$ =158,399.96
  - Standard Deviation = 398
- **Undiscounted Case (n = 240):**
  - Variance = 4,636,840.00 = $4{,}636{,}840 \cdot 2.545$=11,800,757.8
  - Standard Deviation = 3,435.22
- **Infinite Horizon Case:**
  - Variance = $64{,}381.19 \cdot 2.545$=163,850.12
  - Standard Deviation = 404.78

The standard deviations are large relative to the mean firm value of 636.5 (12.73/0.02), highlighting both significant downside risks and substantial upside potential.

In Figures 7a and 7b, the variances of the present value distributions for Blackjack hold percentages are shown for n=240 (20 years) and n=120 (10 years), illustrating the difference between limited and unlimited time horizons. It is evident that, as q increases, the variance for the unlimited time horizon approaches that of the limited time horizon. Practically, the infinite horizon formula provides a quick means to assess risk without needing the exact number of time periods, making it ideal for evaluating long-term stochastic cash flows.

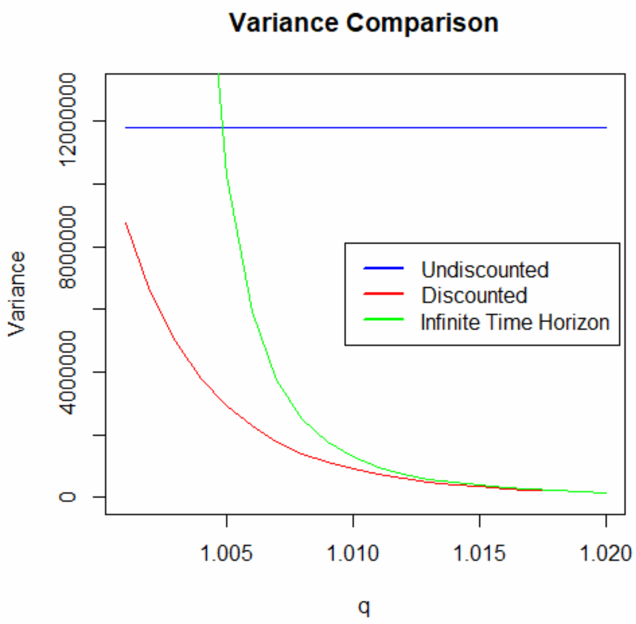


Fig. 7a: Variance comparison depending on q (n=240, $\sigma^2 = 2.545$)

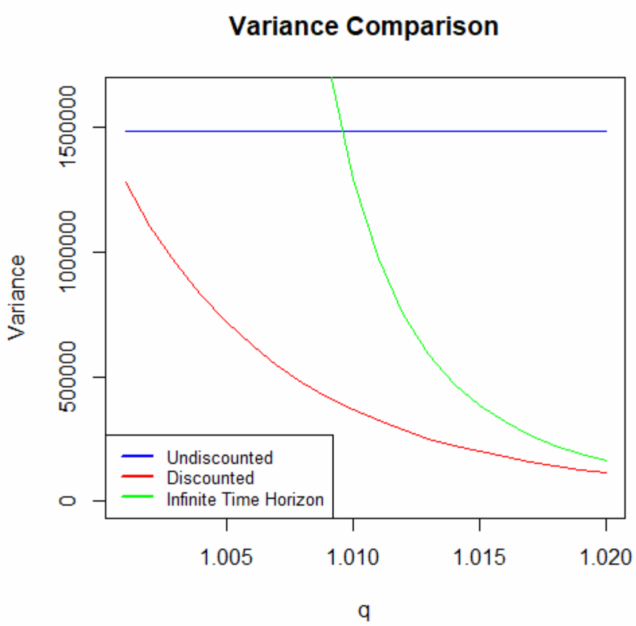


Fig. 7b: Variance comparison depending on q (n=120, $\sigma^2 = 2.545$)

While assuming that hold ratios follow a random walk provides a useful initial approximation for estimating variance and understanding the dynamics of firm values, it may lead to an overestimation of risk. More complex ARIMA models, such as ARIMA(0,1,1)[5], can capture underlying autoregressive patterns and potentially yield a more accurate assessment of variance. Therefore, exploring these advanced modeling approaches is crucial for refining risk evaluations and enhancing the overall analysis of the firm's financial performance.

To estimate the variance σ^2 from the differences $u_t = u_{t-1} + \varepsilon_t$, we can calculate the variance of these differences directly, as they represent the random shocks (or increments) in the random walk. Since each difference corresponds $u_t - u_{t-1}$ to the disturbance ε_t in the random walk model, the variance σ^2 can be estimated by:

1. First, compute the first differences of the cash flow series:

   $$\Delta u_t = u_t - u_{t-1}$$

2. Then, calculate the sample variance of these differences:

   $$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{2}^{n} \left(\Delta u_t - \bar{\Delta} u_t \right)^2$$

   where $\bar{\Delta} u_t$ is the mean of the differences, and n is the number of observations.

This approach provides an estimate of σ^2, the variance of the random shocks driving the random walk, without the need for fitting an ARIMA(0,1,0) model. It leverages the assumption that the increments are independent and identically distributed, allowing for a straightforward calculation based on observed cash flow changes.

Conclusion

This paper presents a stochastic framework for partial firm valuation, analyzing casino hold percentages using Random Walk and ARMA(1,1) models. By modeling cash flows as stochastic processes, we derived variance formulas for discounted and undiscounted cash flows, highlighting the importance of considering random fluctuations and temporal dependencies in risk assessments.

The Random Walk approach offers a simple, practical method for valuation, while ARMA models provide greater accuracy by capturing autocorrelations. Our analysis of real-world casino data from 2004 to 2024 demonstrates that neglecting these dependencies can lead to an underestimation of risk.

In summary, this research emphasizes the value of advanced modeling techniques in firm valuation, offering practical insights for industries with volatile cash flows. Furthermore, further research employing more complex ARIMA models is necessary for a more nuanced

[5] An ARIMA(0,1,1) model with MA(1) parameter $\theta = -0.857$ and AIC = 799.44 results in a significantly reduced standard deviation of the firm value, yielding 72.

calculation of variance in firm valuation, which will enhance risk assessments and decision-making processes in dynamic environments

References

Chatfield, C. (2004): The Analysis of Time Series: An Introduction (6th Edition). CRC Press.

Gerber, H.G. (1971): The discounted Central Limit Theorem and its Berry Esseén Analogue, Ann. Math. Stat. 42, 389-392.

Hyndman, R.J.; Athanasopoulos G. (2018): Forecasting, principles and practice, 2nd edition, OTexts, Melbourne, Australia. OTexts.com/fpp2.

Hyndman, R. et al. (2020): Forecast, Forecasting functions for time series and linear models. R package version 8.13, https,//pkg.robjhyndman.com/forecast/.

Jöckel, K.-H.; Pflaumer, P. (1981a): Die Vorhersage des Goldpreises mit dem Box-Jenkins-Verfahren/Forecasting Monthly Gold Prices with the Box-Jenkins Approach, Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik), vol. 196 6, 481-502.

Jöckel, K.-H.; Pflaumer, P. (1981b): Demographische Anwendungen neuerer Zeitreihenverfahren, Zeitschrift für Bevölkerungswissenschaft, 7, 519-542.

Jöckel, K.-H.; Pflaumer, P. (2022): Exploring Excess Mortality during the Corona Pandemic with Seasonal ARIMA Models. In K.N. Zafeiris, C.H. Skiadas, et al. (Eds.), Data Analysis and Related Applications, Volume 2: Multivariate, Health and Demographic Data Analysis, Wiley, 303-335.

Jöckel, K.-H.; Pflaumer, P. (2024): Using ARMA Models in Stochastic Enterprise Valuation, SEFBIS Journal, March:1-8, DOI: 10.14267/SEFBIS.2024.02.

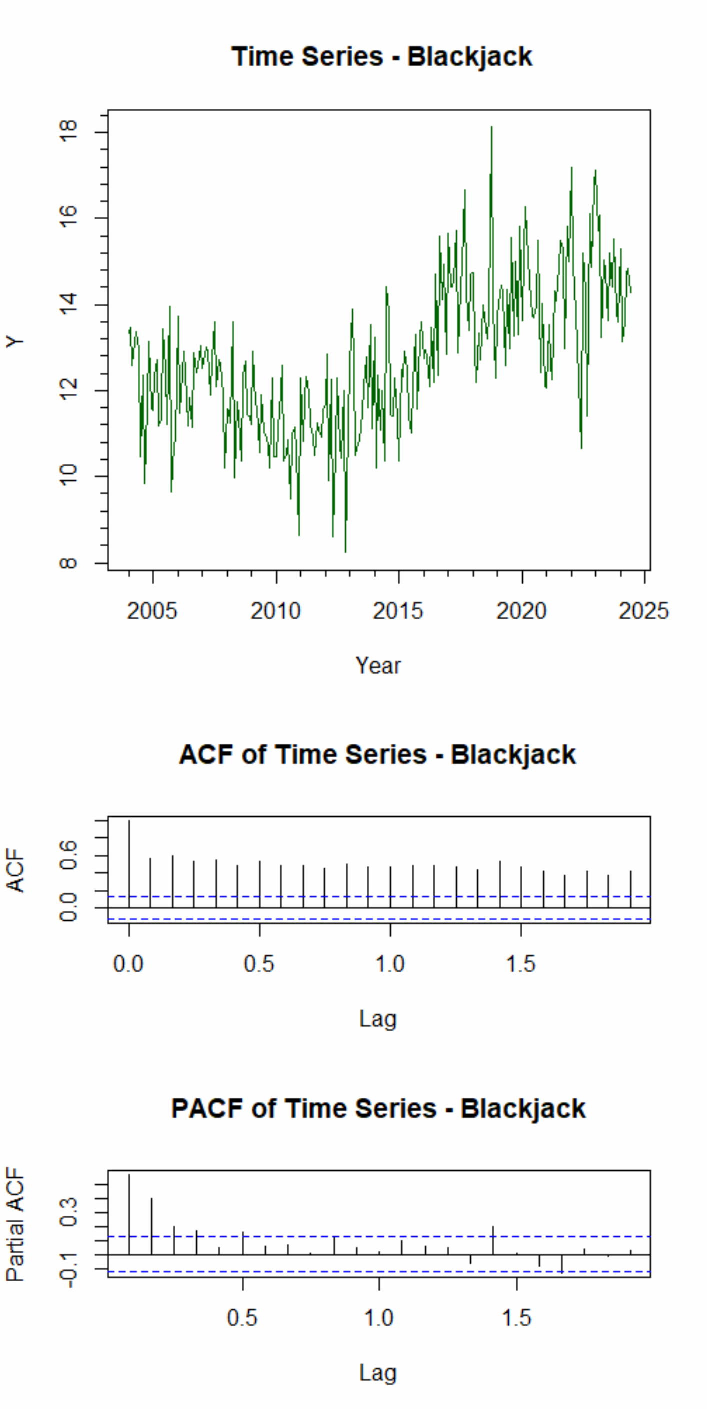
Jöckel, K.-H.;Pflaumer, P. (2025): Present Value Distribution of Random Walks: Analyzing Hold Ratios in Blackjack. *To appear.*

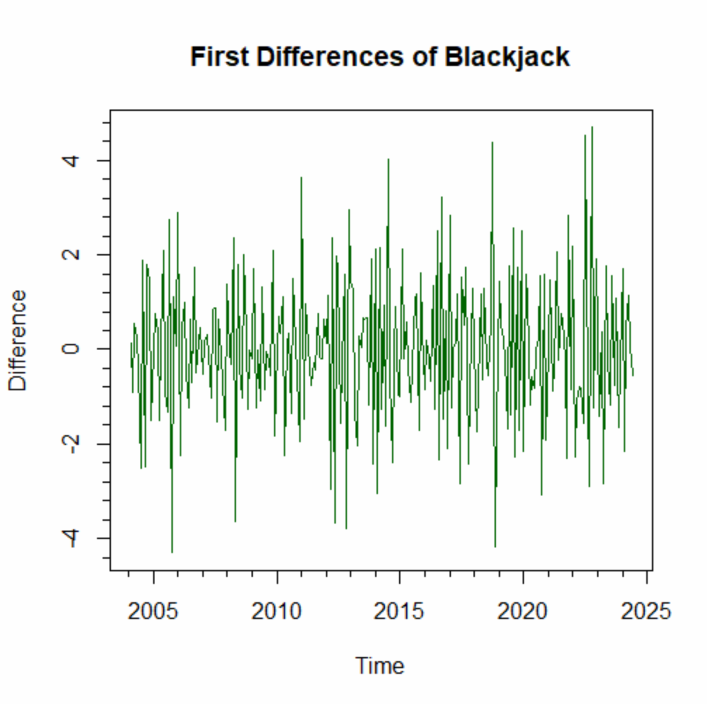
Jöckel, K.-H.; Sendler (1981): A Central Limit Theorem for Generalized Discounting; Mathematische Operationsforschung und Statistik, 12,4, 605-608.

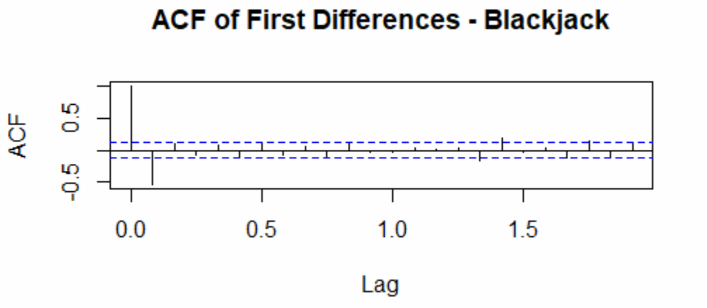
Nau, R. (2020): Statistical forecasting, notes on regression and time series analysis, Fuqua School of Business, Duke University, Durham, NC, https,//people.duke.edu/~rnau/411arim.htm.

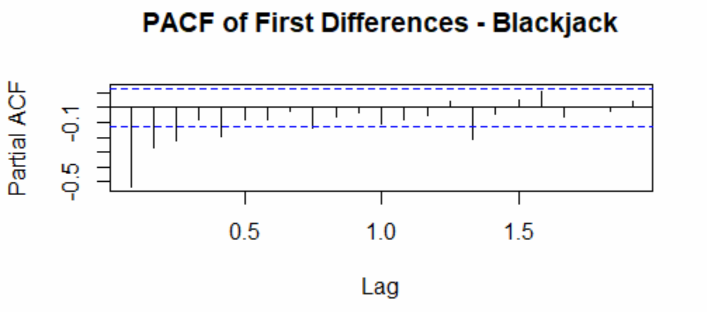
R Core (2020): R, A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https,//www.R-project.org/.

Appendix: Analysis of Blackjack Time Series









Series: Y_imputed
ARIMA(0,1,1)
Coefficients:
ma1
-0.8507
s.e. 0.0394
sigma^2 = 1.503: log likelihood = -397.72
AIC=799.44 AICc=799.49 BIC=806.44Call:

ARIMA(0,1,0)
arima(x = Y_imputed, order = c(0, 1, 0))
sigma^2 estimated **as 2.545**: log likelihood = -462.08, **aic = 926.16**