Volume 39(2), pp. 159–178
http://orion.journals.ac.za

ORiON
ISSN 0259–191X (print)
ISSN 2224–0004 (online)

A portfolio selection problem arising from real-estate investments

S Butler*, KD Searle† and PC Malan*

Received: 14 September 2022; Revised: 15 May 2023; Accepted: 15 May 2023

Abstract

Investments have been used as a medium to mitigate the effects of inflation for many years and it is expected that they will be used for many years to come. Not only do investments allow for an opportunity to increase the net worth of a sum of money, they also allow for a source of passive income. With every investment, however, there is an associated risk but, higher risks are typically associated with a more enticing reward. From the broad scope of potential investment opportunities, the primary focus of this paper is to consider investments within the realm of real estate.

Although property investments have the potential of generating a satisfactory return and may generally be considered a safe investment, as with any investment, poor decision making may still results in the loss of capital. A multi-period portfolio selection model may assist a potential investor in determining an optimal investment plan. The multi-period portfolio selection model considers the future value and rental incomes of a set of potential properties, in determining the optimal investment plan over a given time horizon, aimed at maximising the expected net present value of the portfolio. The quality of the investment plan, however, is dependent on the accuracy of the predicted future value and rental income of the properties under consideration. Therefore, in order to determine these future values accurately, a time series forecasting model is proposed. The forecasting model predicts the values for property values and rental income over a selected time horizon, which serves as input to the multi-period portfolio selection model.

In short, the goal of this paper is to apply suitable time series forecasting methods in order to generate predicted values with an acceptable accuracy, to serve as input for a multi-period portfolio selection model to determine an optimal investment plan as output.

Key words: Multi-period portfolio selection; Forecasting; Property investments.

*Stellenbosch Unit for Operations Research in Engineering, Department of Industrial Engineering, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa

†School of Mathematics and Maxwell Institute for Mathematical Sciences, University of Edinburgh, James Clerk Maxwell Building, Kings Buildings, Mayfield Road, Edinburgh, EH9 3FD, United Kingdom, email: kitsearle@gmail.com

http://dx.doi.org/10.5784/39-2-001

1 Introduction

Inflation is defined as the sustained increase in the average price paid for a given "basket" of goods [1]. Inflation has affected the modern world in various forms for many years and it is expected that it will continue to affect the world in the future. Although inflation rates fluctuate over time, as illustrated graphically in Figure 1, and may vary between countries, it is a reality for every civilian [8]. A consequence of inflation is that capital, if not invested, will depreciate over time [6]. For this reason, capital is invested with the primary objective of earning a *Rate of Return* (ROR) that is greater than that of inflation. If this goal is achieved, the value of the total sum of money invested at the present time will be worth more in the future as a result of its earning capacity being greater than that of inflation [8]. Achieving a ROR that is consistently greater than inflation is not typically an easy task, and requires intelligent investment decisions.

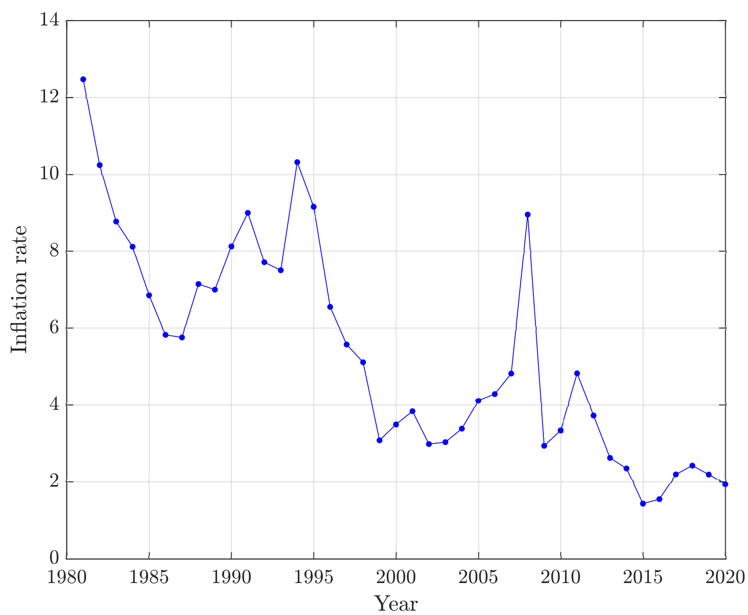


Figure 1: A graphical illustration of the fluctuations in inflation over time according to the World Bank [7].

In an attempt to ensure that investments generate a ROR, that is greater than inflation, investment managers turn to different investment instruments and employ variety of investment strategies. The main investment instruments include cash, bonds, stocks, and property. Each investment comes with a certain level of risk and it is often expected that lower risk investments may result in a lower average ROR in the long term. Many private individuals have been exposed to the property market by purchasing a residential property for the purpose of private occupation. It should be noted that these properties do not generate rental income but instead incur bond repayments and as a result they do not have an annual earning capacity and therefore are not always considered as an investment [9]. While an investment in real-estate is often considered a low-risk investment, real-estate can in fact be a better alternative to stocks as it offers lower-risks, good returns, and allows for greater diversification [3]. An example of how lucrative real-estate investments can be is the business model of one of the world's largest fast-food outlets, McDonalds. Where in 2018 their real-estate portfolio realised more revenue than that of their food sales [4]. Real-estate investments typically yield a lower ROR than what is achieved on the stock market yet, generally they obtain a higher ROR than that of bonds, while offering a lower risk to return ratio [17].

An investment in real-estate is often considered a more stable investment than that of stocks as it can provide both a constant capital inflow in the form of rental income and growth in value which is generally tied to inflation. These characteristics of property investments make it a unique investment opportunity that offers benefits that other investment options lack [17]. Moreover, an investment in real-estate provides the opportunity to leverage an investment through loans.

As a result of the low levels of liquidity and high initial capital requirements associated with an investment in real-estate, there is often a substantial risk of suffering sever finical loses as a result of a bad investment. Therefore, the ability to determine the optimal portion of the initial capital to invest in each investment opportunity will simplify the investment decision making process and generate greater returns due to the elimination of human error.

Capital budgeting is the process of evaluating different investment opportunities to determine their financial viability and devise an investment strategy to maximise the *Net Present Value* (NPV) over some time horizon [20]. This would allow an investor to determine how much to invest in each potential investment opportunity at each point in time. Additionally, it may aid in determining the loan value to take out in each time period. Finally, minimum diversification rules may be implemented within the capital budgeting problem to ensure the investor has an adequate level of diversification.

The solution to a capital budgeting problem is, however, only as accurate as the estimated cash flows and future asset values that are employed as input to the model. This brings about the need for forecasting models that can predict the future value and annual cash flow for a property in a specific demographic region. Forecasting is defined as the rational prediction of future outcomes based on linked current and past information [19]. Given enough historical data on property prices in an area, various forecasting methods may be employed to predict the future value of a property located in that area. Utilising an accurate forecasting model as an input to a capital budgeting problem allows investors to confidently invest in certain properties, in an attempt to maximise their ROR.

In summary, investments in real-estate are utilised throughout the world to ensure that the value of capital is not reduced as a result of inflation. While deriving a capital budgeting model can assist in determining the optimal allocation of capital amongst various different investment opportunities, the solution of this model is only as accurate as the model inputs. With the ever-developing field of data science, an increasing amount of available data, and continual improvements to the analysis of data, these developments can be employed within the field of real-estate investment to improve the model inputs for a capital budgeting model. A framework, called the *investment solution process* is proposed in this paper for determining the optimal allocation of starting capital to maximise the NPV of a given time horizon.

2 Literature review

Portfolio optimisation is concerned with the selection of assets over a time period while considering the investor's risk-return status. In 1952 Markowitz [10] proposed the so-called mean-variance model which was the first attempt at modeling portfolio optimisation and resulted in increased attention in portfolio optimisation by both academics and liability professionals. The mean-variance approach considers how investors should construct optimal portfolios by taking into account the risk associated with market volatility while maximising the expected returns [2].

There have been many reformulations of the model developed by Markowitz, in this section, however, the focus is placed solely on the case in which the investor's objective is to maximise the expected return and is subject to risk constraints [15]. This results in the following optimisation problem,

$$\max_{\boldsymbol{x}} \quad z = \boldsymbol{r}^T \boldsymbol{x}, \tag{1}$$

$$\text{s.t} \quad \boldsymbol{x}^T \boldsymbol{R} \boldsymbol{x} \leq \sigma^2, \tag{2}$$

$$\sum_{i=1}^{n} x_i = 1, \tag{3}$$

$$x_i \in [0, 1] \quad \forall i \in \{1, 2, ..., n\}, \tag{4}$$

where $\boldsymbol{x} = [x_1, \ldots, x_n]^T$ is fraction of capital invested in asset $i \in \{1, \ldots, n\}$. The vector $\boldsymbol{r} \in \mathbb{R}^n$ containing as its i-th component the expected return of investment i, and a variance covariance matrix of $\boldsymbol{R} \in \mathbb{R}^{n \times n}$ such that $\boldsymbol{R}_{i,j} = cov(r_i, r_j)$ for all $i, j \in \{1, \ldots, n\}$. The constraint in (2), therefore limits the risk of the investment, so that it does not exceed the maximum acceptable risk of the investor, denoted by, σ^2 and where the objective function is to maximise the expected return.

The mean-variance framework has been a popular framework for portfolio optimisation problems since the 1960's [15]. This includes employing results from dynamic programming and Bellman's principle of optimality to determine an individuals optimal consumption policy [13]. Mossin [12] considered the topic of employing results from dynamic programming to determine the optimal policies for both single-period and multi-period optimisation problems. Dynamic programming was also employed by Samuelson [16] with the objective of determining an optimal lifetime consumption of an individuals lifetime income for investment and in investment policy. There have been considerable developments in the field of portfolio optimisation models over the last few decades, several of these have been limited as a result of the slow advances in available solution methodologies and the computational capabilities of computers [15]. The recent advancement in solvers, processor speeds and an increase in memory capabilities has allowed for larger and more realistic models to be solved [15].

While these advancements have been made, the financial markets have become more complex which means that a realistic model must take into account numerous different considerations, rendering the situation more complex [15]. Milhomem & Dantas [11] conducted a review of papers that used exact methods to solve the portfolio optimisation problem which equated to a total 18 papers. This shows strong evidence that these methods are used for the selection of the best assets. According to the study 23.8% of papers used heuristics methods, 38.09% considered a single-period problem, 47.61% address multi-period problems, and lastly 71.41% of the papers have more than one objective function [11]. This shows that despite the evolution of new techniques for the solving of optimisation problems, many investors and researchers prefer exact techniques due to their ability to find the global optimal [11]. Despite the level of difficulty associated with solving portfolio optimiasation problems by employing exact techniques, they always return the optimal solution, making them attractive for investors as a whole and especially for conservative investors [11].

3 The investment solution process

In this section the process of combining a capital budgeting problem with a time series forecasting model in a system called the *investment solution process* is described. For each individual module of the investment solution process, data is taken as input and a desired output is produced which serves as input for the following module. This investment solution process is represented graphically by means of a process flow diagram in Figure 2.

The investment solution process begins with the data source which contains data pertaining to historical time series value and rental income for properties under consideration. This data in addition to a number of user inputs, serve as input to the subsequent modules and are used throughout the investment solution process. The data cleaning and processing module takes as input raw data and returns a clean dataset. Once the data is in a clean and usable format, it may be split into two subsets, namely the testing set and the training set for various test/train ratios. The forecasting module takes as input the time series value and rental training sets and produces as output the predicted value and rental datasets for various forecasting methods and at the different test/train ratios. In the evaluation module, the predicted property values and rental income are compared to the associated test sets. The performance of the various forecasting models, in conjunction with the different train/test splits, are then calculated in terms of a *Mean Absolute Percentage Error* (MAPE) score, before being presented to the user. The user then selects a forecasting method, together with the required train/test split. The resulting output of the forecasting module serves as input to the capital budgeting module in the form of forecasted property values and rental values across the different properties. The capital budgeting module then returns the optimal investment plan as well as the NPV of the portfolio to the user. In the remainder of this paper, each module of the investment solution process is discussed in more detail.

3.1 Property data source

The datasets which are required for the implementation of the investment solution process may be sourced from any data source, however, the format of the data must adhere to certain requirements. The first being that the property data must contain two time series data sets pertaining to the value of each property together with their associated rental income which are called the value and rental datasets, respectively. The minimum

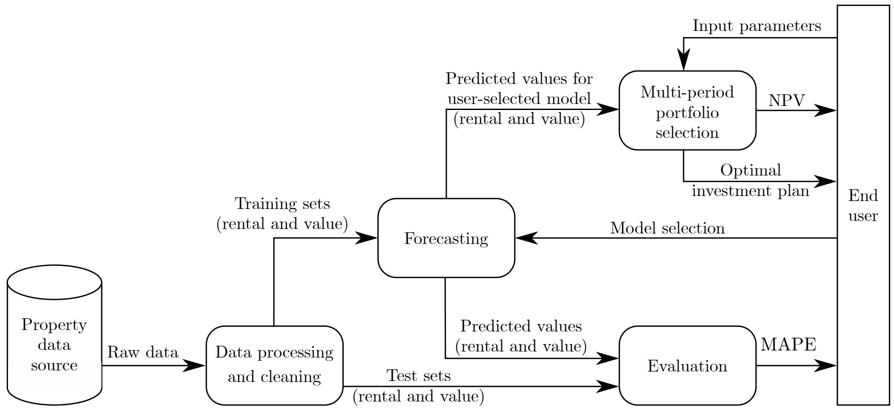


Figure 2: A process flow diagram graphically illustrating the investment solution process.

requirements for the attributes within the datasets are demonstrated in Tables 1 and 2, respectively.

Table 1: A table demonstrating the layout of the median sale price per property for each month after cleaning and manipulation.

Region code	Property	t_1	t_2	...	t_T
1	Property 1	370 000	362 500	...	515 000
2	Property 2	18 700	174 000	...	295 000
⋮	⋮	⋮	⋮	⋱	⋮
N	Property N	250 000	236 872	...	470 000

Table 2: A table demonstrating the layout of the median rental income per property for each month after cleaning and manipulation.

Region code	Property	t_1	t_2	...	t_T
1	Property 1	2 266	2 295	...	2 549
2	Property 2	1 464	1 477	...	7 140
⋮	⋮	⋮	⋮	⋱	⋮
N	Property N	1 200	1 238	...	1 629

As a requirement for the capital budgeting model within this solution methodology, the rows within the datasets must contain the various properties and the columns must be filled with their associated prices and rental income. Furthermore, it is noted that the datasets may have additional attributes, a differing number of properties or have different time frames. This is not an issue within the investment solution process as the data processing and cleaning module is capable of handling these differences.

3.2 Data processing and cleaning module

It is important that the two datasets contain the same attributes, the same number of unique regions and have the same number of time series instances. To this end, it is a requirement for attributes to be removed if the datasets do not align. Furthermore, if the two datasets do not contain the time instances, then the time instances that are not present in both datasets are subsequently removed.

Finally, the region code is matched for each dataset and the properties that are not present in both datasets are subsequently removed. In other words, each unique property in the value dataset is required to have a unique (matching) property in the rental dataset. The completion of these steps results in two datasets with matching properties and the same number of discrete time instances. Thereafter, the two datasets are split into training and testing datasets, through the use of three different train/test splits. The first train/test split is formed with the training dataset consisting of the first 50% of the total discrete time instances and the testing dataset consisting of the remaining 50% of discrete time instances. The second train/test split is selected so that the training dataset constitutes the first 63% of the discrete time instances and the testing dataset constitutes the remaining 37% of the discrete time instances. Finally, the third train/test split is selected such that the training dataset constitutes the first 75% of the discrete time instances and the testing dataset constitutes the remaining 25% of the discrete time instances. At this point we note that the appropriate train/test split will depend on the volume of available data and may differ for each user or case study.

3.3 Time series forecasting module

The time series forecasting module takes as input the clean and processed data *(i.e. the partial output of the data processing and cleaning module)*. More specifically, this model utilises the training sets for both the value and rental incomes of each property. The data is analysed to ensure it is in the correct format, before checking for any trends and seasonality present within the dataset. Different time series forecasting methods are employed, such as the naive forecaster, exponential smoothing and *Auto Regressive Integrated Moving Average* (ARIMA) to forecast both the median sale price and the median rental income for each property. The resulting accuracy of the forecasts are subsequently compared, in order to select the appropriate forecasting method. To this end, the output of the time series forecasting module includes the predicted values for the value and rental datasets resulting from the different forecasting methods and train/test splits. The output of this module then serves as input to the evaluation module and ultimately the multi-period portfolio selection module, after the end user has determined which method to employ together with the required train/test split.

3.4 Evaluation module

The evaluation module takes as input the predicted datasets for the value and rental income, in addition to the test datasets of the value and rental income from the data cleaning and processing module. The two inputs are then used to create a comparison matrix which enables for the accuracy of the predicted dataset to be determined, by means of the MAPE of the forecast. The MAPE of the various forecasting methods and train/test splits are then compared, allowing for the selection of the best performing method and split. Once the final forecasting technique and train and test split is selected the resulting MAPEs of the value and rental income predictions, are provided to the user as an indication of the accuracy of the predictions. The data for this selected forecasting technique then serves as input for the multi-period portfolio selection module which was derived in §3.5.

3.5 The multi-period portfolio selection module

The multi-period portfolio selection module is the final module in the investment solution process. The output of this module is presented to the end user in the form of an optimal investment plan. The rest of this section is devoted to the derivation of the multi-period portfolio selection problem which takes the form of a single-objective *Mixed-Integer Linear Program* (MILP). The objective of this model is to maximise the NPV of a given initial investment, by selecting the optimal investment plan, while adhering to certain constraints. The model output aims to assist the decision maker as to which properties should be bought and sold at discrete time intervals, so as to maximise the NPV over a determined time horizon. This section opens in §3.5.1 with a description of the model parameters. Thereafter, a detailed description of the decision variables used within the model and their importance is provided §3.5.2. The models constraints are then defined in §3.5.3 and finally the objective function is defined and discussed in §3.5.4.

3.5.1 Model parameters

Let $\mathcal{P} = \{1, \ldots, P\}$ denote the set of properties available for purchase and let $\mathcal{T} = \{1, \ldots, T\}$ denote the set of discrete decision time instances. Furthermore, let I denote the initial capital available for investment and let r denote the interest rate that represents the earning potential of I if it is invested in a money market. Let M be a positive integer denoting the minimum diversification requirement, stated by the user, which defines the minimum number of properties that must be owned at any time $t \in \mathcal{T}$. Moreover, let L be a positive real-number that denotes the maximum amount of money that can be borrowed at any particular time $t \in \mathcal{T}$. Finally, let b and i be real numbers denoting the interest rate at which money may be borrowed and the inflation rate, respectively.

Let the value of each property $i \in \mathcal{P}$ at any time $t \in \mathcal{T}$ be denoted by the real number $V_{i,t}$. Therefore, the purchasing price, denoted by $P_{i,t}$ of each property $i \in \mathcal{P}$ at any time $t \in \mathcal{T}$ can be determined as a function of the value of the property. To this end, let $AC_{i,t}$, denote the acquisition costs associated with purchasing of property $i \in \mathcal{P}$ at time $t \in \mathcal{T}$. The purchasing price of a property $i \in \mathcal{P}$ in time period $t \in \mathcal{T}$ may be defined as

$$P_{i,t} = V_{i,t} + AC_{i,t}.$$

Similarly, parameter $V_{i,t}$ is again considered when determining the selling price, denoted by $S_{i,t}$, of each property $i \in \mathcal{P}$ at any time $t \in \mathcal{T}$. Therefore, let $SC_{i,t}$ denote the selling costs associated with property $i \in \mathcal{P}$ at any time $t \in \mathcal{T}$. The selling price of property $i \in \mathcal{P}$ at any time period $t \in \mathcal{T}$ may be defined as

$$S_{i,t} = V_{i,t} - SC_{i,t}.$$

The return generated by an owned property is not only determined by the difference in its purchasing price $P_{i,t}$ and its selling price $S_{i,t}$, but also by the property's earning capacity through rental income. Therefore, let $R_{i,t}$ be the *Net Cash Flow* (NCF) of rental income generated by a property $i \in \mathcal{P}$ and at any time $t \in \mathcal{T}$. That is, the rental NCF accounts for the rental income after deducting all running costs associated with owning property $i \in \mathcal{P}$ in time period $t \in \mathcal{T}$.

3.5.2 Decision variables

Five decision variables, contained in the set $\mathcal{D}$, are employed within this model. The first variable models the situation where property $i \in \mathcal{P}$ is purchased at time $t \in \mathcal{T}$. Therefore, let

$$x_{i,t} = \begin{cases} 1 & \text{if property } i \in \mathcal{P} \text{ is purchased at time } t \in \mathcal{T}, \\ 0 & \text{otherwise,} \end{cases}$$

where the assumption is made that no properties may be purchased in the final time period, that is, $x_{i,T} = 0 \ \forall i \in \mathcal{P}$. Similarly, a second variable is defined to model the situation where a property $i \in \mathcal{P}$ is sold at time $t \in \mathcal{T}$. Therefore, let

$$y_{i,t} = \begin{cases} 1 & \text{if property } i \in \mathcal{P} \text{ is sold at time } t \in \mathcal{T}, \\ 0 & \text{otherwise,} \end{cases}$$

where the assumption is made that no properties may be sold in the first time period, that is, $y_{i,1} = 0 \ \forall i \in \mathcal{P}$.

The third decision variable, which is in fact an auxiliary variable, keeps track of the properties in set $\mathcal{P}$ that are currently owned during time period $t \in \mathcal{T}$, and is determined by

$$z_{i,t} = \sum_{n=0}^{t} x_{i,n} - \sum_{n=0}^{t} y_{i,n} \quad i \in \mathcal{P},\ t \in \mathcal{T}.$$

The fourth decision variable denoted by C_t is the amount of capital left on hand during time period $t \in \mathcal{T}$. Finally, let $B_{t,p}$ denote the amount of money borrowed at time $t \in \mathcal{T}$ and paid back at time period $p \in \mathcal{T} \ \forall \ t < p$.

3.5.3 Model constraints

The model consists of a total of nine constraints, all with the purpose of ensuring the model output adheres to certain budgeting requirements. The first constraint

$$I + \sum_{p=2}^{T} B_{1,p} = \sum_{i=1}^{P} (P_{i,1} x_{i,1}) + C_1$$

ensures the total available funds are allocated to either a property investment or to the money market in the first time instant. Thereafter, a recursive constraint is derived to ensure that the total capital is conserved in each time period. That is, the total monetary inflow for time period $t \in \mathcal{T} \setminus \{1, T\}$ is equal to the total monetary outflow for the time period $t \in \mathcal{T} \setminus \{1, T\}$. Specifically the inflows include receiving income from the money market, rental income from the properties owned in the previous period, selling a property or taking a loan. Whereas the outflows are a result of expenditure on paying back a loan, purchasing properties or leaving money in the money market. This may be expressed mathematically as

$$C_{t-1}(1+r) + \sum_{p=t+1}^{T} B_{t,p} + \sum_{i=1}^{P}(S_{i,t}y_{i,t}) + \sum_{i=1}^{P}(R_{i,t-1}z_{i,t-1})$$
$$= \sum_{i=1}^{P}(P_{i,t}x_{i,t}) + C_t + \sum_{n=1}^{t} B_{n,t}(1+b)^{(t-n)}, \quad \forall i \in \mathcal{P},\ t \in \mathcal{T} \setminus \{1, T\}.$$

Additionally the constraint

$$y_{i,n} \leq \sum_{t=0}^{n} x_{i,t-1}, \quad \forall i \in \mathcal{P},\ n \in \mathcal{T}$$

is introduced to ensure that a property is not sold if it had not yet been purchased by the previous time period. Furthermore, in order to ensure the same property is not bought or sold more than once, the constraints

$$\sum_{t=1}^{T} x_{i,t} \leq 1 \quad \forall i \in \mathcal{P}$$

and

$$\sum_{t=1}^{T} y_{i,t} \leq 1 \quad \forall i \in \mathcal{P}$$

are introduced.

Furthermore, a constraint set is employed to ensure that the solution adheres to the diversification parameter M, set by the user. The constraints

$$\sum_{i=1}^{P} z_{i,t} \geq M \quad \forall t \in \mathcal{T} \setminus \{T\}$$

ensure that there are never less than M properties owned in any time period, such that the user's specification is satisfied. The constraints

$$\sum_{p=t+1}^{T} B_{t,p} \leq L \quad \forall t \in \mathcal{T} \setminus \{T\}$$

are introduced to ensure that the amount of money borrowed at each time period $t \in \mathcal{T}$ does not exceed the specified maximum amount that may be borrowed.

Finally the domain constraints

$$C_t \geq 0 \quad \forall t \in \mathcal{T} \setminus \{T\}$$

ensures that the cash flow C_t is non-negative, and the domain constraints

$$B_{t,p} \geq 0 \quad \forall p \in \mathcal{T} \setminus \{1\},\ t \in \mathcal{T} \setminus \{T\}$$

ensures that the amount of money borrowed $B_{t,p}$ is also non-negative.

3.5.4 Model objective function

The goal of the capital budgeting model is to maximise the NPV of the cash on hand at the end of the investment time period T, expressed mathematically as

$$\max_{\mathcal{D}} J = \frac{\sum_{i=1}^{P}(S_{i,T}y_{i,T}) + C_{T-1}(1+r) + (R_{i,T-1}z_{i,T-1}) - \sum_{t=0}^{T-1} B_{t,T}(1+b)^{(T-t)}}{(1+r_i)^T}.$$

3.5.5 Model Output

The investment solution process output provided to the user is that of the optimal investment plan, which is the solution of the multi-period portfolio selection model. It is important that this output is presented in a user friendly manner so that the output is easy to interpret. To this end, the output is presented by creating a table consisting of all the properties as row entries and the time instances as column entries, whereby the table is highlighted when a specific property is owned in a particular time period. This allows the user to easily view when properties are purchased and sold, throughout the investment time horizon.

4 Case Study

This section has the objective of implementing the investment solution process discussed in §3.1 to a real life case study. This is accomplished through the utilisation of a real-life dataset and its implementation within the various modules. This will allow for a comparison of the model's performance to performances of other investment opportunities. The implementation of this case study will be discussed per module within the investment solution process.

4.1 Property data Source

The datasets which are used for the case study are sourced from the Zillow group [21]. The Zillow group is an American based company that offers customers an on-demand real-estate experience, with their purpose of re-imagining the real-estate industry. More specifically, in addition to their buying, selling and renting services, they also offer access to research data pertaining to the housing industry within the United States of America (USA). Within their research data offering for the USA real-estate market, they have Home Value Index (HVI) datasets available for use by the public. This specific index offers a measure of the typical home values and market changes across different regions in the USA. These measures vary from home values, home value forecasts, rentals, and sale prices. From the aforementioned data offered by Zillow group, certain datasets were identified as suitable for use within this project.

The selected datasets consist of median sale prices and the associated median rental income for demographic areas within the USA. This data is downloaded in the form of two separate datasets, the first dataset contains time series data with respect to the value of the property which is called the monthly sale price dataset. The second containing time series data with respect to the rental income of the properties, called the monthly rental income dataset. Both datasets consist of the median values associated with the sale price and rental income of properties within various demographic regions across the USA. It should be noted that the sale price dataset consists of six attributes and contains time series data dating back to 2008, whereas the rental income dataset consists of four attributes and contains time series data dating back to 2014. Finally, it should be noted that the sale price dataset consists of 96 unique regions, whereas the rental income dataset consists of 101 unique regions.

4.2 Data processing and cleaning module

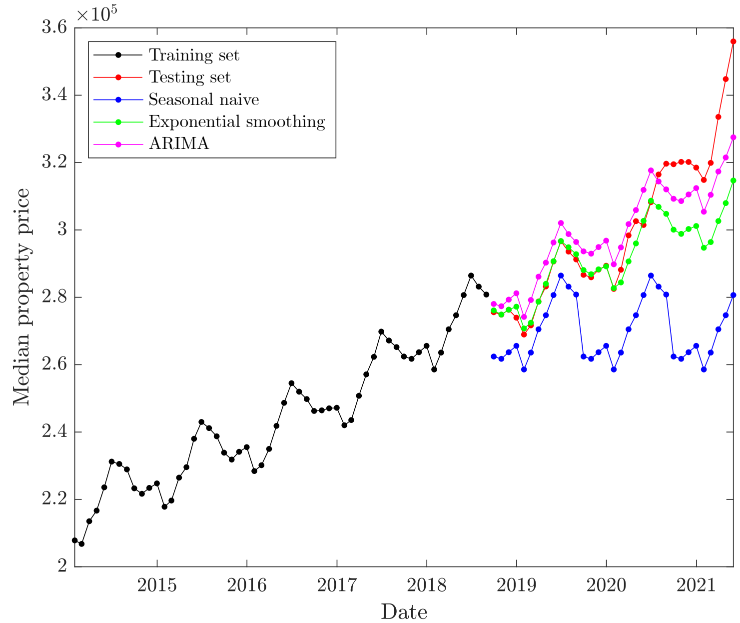
In order to ensure that the datasets meet the requirements discussed in §3.2 several of the attributes are removed to ensure that both datasets only contain the attributes region code, region name and the associated time series data for each region. Thereafter, instances before 2014 in the sale price dataset are removed to ensure the same number of time series instances are contained in each dataset. Finally, the region code is matched for each dataset and the regions that are not present in both datasets are subsequently removed. The completion of these steps result in two datasets with matching regions and the same number of discrete time instances. Thereafter, the two datasets are then split into the training and testing datasets, through the use of the three train/test splits discussed in §3.2.

4.3 Time series forecasting module

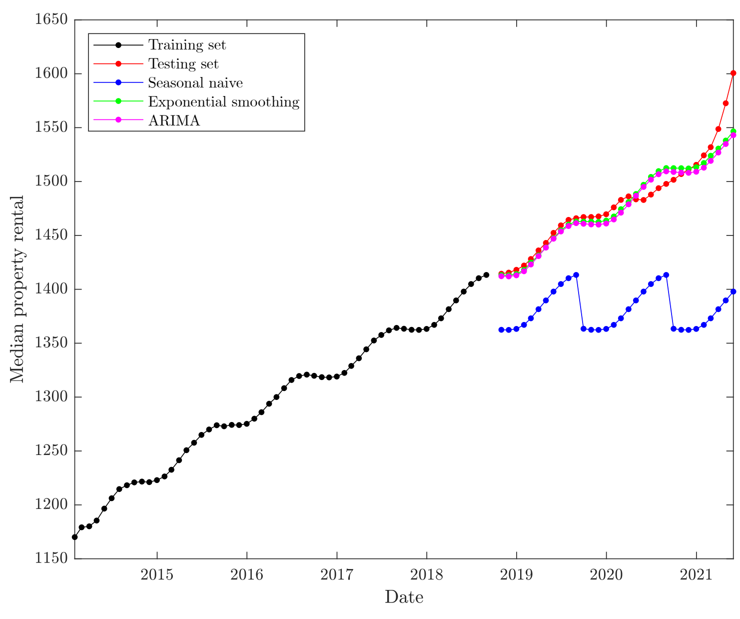
Following the cleaning of the dataset in §4.2, from the above discussed investment solution process, this data is then used as input into the time series forecasting module. The three time series forecasting techniques which were mentioned in §3.3 are applied to the datasets.

The mean values for the median sale price and rental income are first implemented into each forecasting technique to get an idea of the accuracy. This analysis is done through plotting the predicted values against the actual values in an attempt to visualise the accuracy of each forecasting technique. A visual representation of the accuracy for each forecasting technique for both the monthly property price dataset and the monthly rental income dataset is illustrated in Figures 3(a) and 3(b), respectively.

Through this analysis it is evident that the exponential smoothing and ARIMA forecasting methods perform much better than that of the seasonal naive method. It is, of course, not possible to determine which of the two are more accurate, by merely inspecting their corresponding graphs. To determine the exact accuracy for each technique the MAPE is used to determine the prediction accuracy for each property region for each forecasting method. The results are tabulated in Table 3 which shows the different MAPE values for each time series forecasting technique across the various train/test splits. Through the analysis of this table, it is seen that the combination of the 63%/37% train test split along with the exponential smoothing forecasting method produces the most accurate forecasts in terms of the trade off between accuracy and train/test split. This combination is therefore employed to forecast the median property price and median rental income for each region.



(a) Value predictions



(b) Rental income predictions

Figure 3: A line graph representing the predicted values for both the property value and the rental values of the seasonal naive, exponential smoothing and ARIMA forecasting methods.

Table 3: A table showing the MAPE values of each forecasting technique for both the value and rental predictions.

	50%/50%		63%/37%		75%/25%	
	Value MAPE	Rental MAPE	Value MAPE	Rental MAPE	Value MAPE	Rental MAPE
Naive forecast	13.88%	8.45%	11.24%	7.20%	10.59%	5.53%
Exponential smoothing forecast	5.41%	2.41%	4.93%	1.95%	4.90%	1.86%
ARIMA forecast	7.31%	2.88%	6.59%	2.23%	5.19%	1.94%

4.4 Multi period portfolio selection module

As discussed in §3.5, the model requires the predicted property and rental values of the forecasting module as input in addition to a number of inputs required from the user. For this specific case study, these additional user inputs are tabulated in Table 4.

Table 4: The values of the model parameters used for the implementation of the solution investment process.

Parameter	**Value**
Initial capital available	$1 000 000
Diversification constraint (minimum number of properties)	2
Monthly maximum loan amount	$50 000
Borrowing rate	3.21%
Interest rate	5.4%

Furthermore, the case study dataset does not contain any information pertaining to the costs associated with the purchasing and selling of the properties. To account for this an average percentage cost associated with the purchasing of a property in the USA is determined and an average percentage cost associated with selling of a property is determined. This is to allow for the costs associated with purchasing and selling a property to be determined, which is required as input for the derived model in §3.5. As of January 2020, the additional costs that are incurred when a property is transferred, are shown in Table 5. It can be seen that the total percentage costs which are paid by the buyer range from 1.20% to 2.50%, whereas the total percentage costs incurred by the seller range from 7.50% to 8.425%. To account for the worst case scenario, the purchasing costs is taken to be 2.5% of the value of the property whereas the selling cost is taken to be 8.425% of the value of the property.

4.5 Discussion of results

For this case study discussion of the results, the optimal solution to the multi-period portfolio selection is briefly discussed. With regards to the case study dataset that is

Table 5: A table showing the different percentage of costs incurred by the seller and buyer of a property [5].

Transaction cost	Cost percentage to value of property	Who incurs the cost
Title search and insurance	0.5%–1.0%	Buyer
Recording fee	0.2%–0.5%	Buyer
Legal fees	0.5%–1.0%	Buyer
Legal fees	0.5%–1.0%	Seller
Real property transfer tax	1.0%–1.425%	Seller
Real estate brokers fee	6.00 %	Seller

used, the optimal output from the multi-period portfolio selection model may be seen in Appendix A:. This output is interpreted to see that the initial regions that are invested in, at the first time instance include region 76 and region 81. Following this, it is seen that once enough capital is acquired, further regions are invested in throughout the time horizon of the investment. In the second last time instance, a total of 13 regions are invested in, before they are all sold in the last time instance. Through this optimal investment plan, the NPV of the portfolio is increased to be worth \$1 636 200. This implies that a profit of \$636 200 is made after recovering the initial investment. When considering the time frame of thirty-three months which is the equivalent of two and three quarter years, there was a resultant annual return of \$231 345 per year which equates to a return of 23.13%. By comparison, the historical annual average stock market return is 10% according to Royal & O' Shea [14]. When comparing this return to that of the property investment it is seen that the model and property investment is able to outperform that of the stock market.

5 Conclusion

In this paper, a multi-period portfolio selection model was proposed as part of the investment solution process to assist potential investors in determining how to distribute their initial investment across available investment opportunities, in the form of an optimal investment plan. This multi-period portfolio selection model, takes as input future predicted property prices and their associated rental income, and provides as output an optimal investment plan in the form of recommendations as to which properties to invest in at each discrete time instance. Furthermore, time series forecasting methods are employed to take a set of data as input and predict the future values for each property as output for use in the multi-period portfolio selection model. Not only does this model advise an investor as to where to invest capital, it also advises when the investor should take out and pay back loans. The benefit of such a model is that it may help an investor maximise the possible return on their investment through an optimal loan strategy in addition to an optimal investment strategy. The model also possesses a degree of individual usability as it takes as input the individual investor's chosen property dataset, their available initial capital, the maximum allowable loan amount, their given time frame and diversification specification. This results in a robust model that can be used by different investors with varying circumstances. Not only does this flexibility allow for different users but it also allows for a single user to test different scenarios to see which is the best option.

Even though the model provides as output an optimal investment plan, it is important for the user to utilise this as a guide rather than as a set investment plan. This is important as it is impossible to predict future property prices with 100% accuracy, and in reality the optimal solution plan can never be 100% accurate. Moreover, market changes caused by external forces can greatly effect the investment plan, and the interest and borrowing rates will typically vary in time whereas they are constant in the context of this model.

When interpreting the results of the model, it was clear that the model output neglects the lack of liquidity that is associated with property ownership, as the model output suggests the purchase and sale of properties regularly. This is a result of the limiting assumption that a buyer will invariably be available whenever a property needs to be sold. This is not the case in the real-world, however, as properties often remain on the market for a significant period of time. Moreover, the model assumes a fixed borrowing and interest rate over the entire investment horizon, which is an unrealistic assumption as these rates vary significantly over time. These varying rates could significantly effect the model output.

References

[1] Blank L & Tarquin A, 2012, *Engineering Economy*, 7th Edition, McGraw Hill, New York (NY).

[2] Bonami P & Lejeune MA, 2009, *An exact solution approach for portfolio optimization problems under stochastic and integer constraints*, Operations Research, **57**, pp. 650–670.

[3] Boykin R, 2021, *Reasons to invest in real estate vs. stocks*, [Online], [Cited: 10 January 2022], Available from `http://www.investopedia.com/investing/reasons-invest-real-estate-vs-stock-market`.

[4] Egan J, 2019, *Under new leadership, will McDonald's grill up a new real estate strategy?*, [Online], [Cited: 10 January 2022], Available from `https://www.wealthmanagement.com/investment/under-new-leadership-will-mcdonald-s-grill-new-real-estate-strategy`.

[5] Global Property Guide, 2020, *U.S. buying costs range from low to moderate*, [Online], [Cited: October 2021], Available from `https://www.globalpropertyguide.com/North-America/United-States/Buying-Guide`.

[6] Hayes M, 2011, *The social history of quantifying inflation: A sociological critique*, Journal of Economic Issues, **45**, pp. 97–112.

[7] International Monetary Fund, 2021, *Inflation, consumer prices (annual %)*, [Online], [Cited: 27 April 2021], Available from `https://data.worldbank.org/indicator/FP.CPI.TOTL.ZGend=2019&name_desc=false&start=1981&view=chart&year=2019`.

[8] López-Díaz, MC, López-Díaz M & and Martínez-Fernández S, 2018, *A stochastic order for the analysis of investments affected by the time value of money*, Insurance: Mathematics and Economics, **83**. pp.75–82.

[9] Luüs C, 2015, *Long term real investment returns*, [Online], [Cited: 10 January 2022], Available from http://www.financialmarketsjournal.co.za/long-term-real-investment-returns.

[10] Markowitz H, 1952, *Portfolio selection*, The Journal of Finance, **7**, pp. 77–91.

[11] Milhomem DA & Dantas MJP, 2019, *Analysis of new approaches used in portfolio optimization: A systematic literature review*, Production, **30**, Manuscript e20190144 (no pae numbers).

[12] Mossin J, 1968, *Optimal multiperiod portfolio policies*, The Journal of Business, **41**, pp. 215–229.

[13] Phelps ES, 1962, *The accumulation of risky capital: A sequential utlity analysis*, Econometrica, **30**, pp. 729–743.

[14] Royal J & O'Shea A, 2021, *What is the average stock market return?*, [Online], [Cited: October 2021], Available from https://www.nerdwallet.com/article/investing/average-stock-market-return.

[15] Sağlam Ü & Benson HY, 2019, *Single and multi-period portfolio optimization with cone constraints and discrete decisions*, Available at SSRN: https://ssrn.com/abstract=2820294orhttp://dx.doi.org/10.2139/ssrn.2820294.

[16] Samuelson P, 1975, *Life time portfolio selection by dynamic stochastic programming*, Stochastic Optimization Models in Finance, **51**, pp. 517–524.

[17] Szumilo N, Wiegelmann T, Laszkiewicz E, Pietrzak MB & Balcerzak AP, 2018, *The real alternative? A comparison of German real estate returns with bonds and stocks*, Journal of Property Investment and Finance, **36**, pp. 19–31.

[18] Svestka J & Huckfeldt J, 1973, *Computational experience with an M-salesman traveling salesman algorithm*, Management Science, **15**, pp. 790–799.

[19] Wilson ID & Paris SD & Ware JA & Jenkins DH, 2018, *Residential property price time series forecasting with neural networks*, Knowledge-Based Systems, **15**, pp. 335–341.

[20] Winston WL, 2004, *Operations Research Applications and Algorithms*, 4^{th} edition, Brooks/Cole, Belmont (CA).

[21] Zillow Group, 2021, *Zillow home value index*, [Online], [Cited: July 2021], Available fromhttps://www.zillow.com/research/data/.

Appendix A: Multi-period portfolio selection model output

The output of the decision variable $z_{i,t}$ of the multi-period portfolio selection module is provided in Figures 4–6.

	Region ID	Region Name	Region Number	2018-09-30 00:00:00	2018-10-31 00:00:00	2018-11-30 00:00:00	2018-12-31 00:00:00	2019-01-31 00:00:00	2019-02-28 00:00:00	2019-03-31 00:00:00	2019-04-30 00:00:00	2019-05-31 00:00:00
0	102001	United States	Region_0									
1	394913	New York, NY	Region_1									
2	753899	Los Angeles-Long Beach-Anaheim, CA	Region_2									
3	394463	Chicago, IL	Region_3									
4	394514	Dallas-Fort Worth, TX	Region_4									
5	394974	Philadelphia, PA	Region_5									
6	394692	Houston, TX	Region_6									
7	395209	Washington, DC	Region_7									
8	394856	Miami-Fort Lauderdale, FL	Region_8									
9	394347	Atlanta, GA	Region_9									
10	394404	Boston, MA	Region_10									
11	395057	San Francisco, CA	Region_11									
12	394532	Detroit, MI	Region_12									
13	395025	Riverside, CA	Region_13									
14	394976	Phoenix, AZ	Region_14									
15	395078	Seattle, WA	Region_15									
16	394865	Minneapolis-St Paul, MN	Region_16									
17	395056	San Diego, CA	Region_17									
18	395121	St. Louis, MO	Region_18									
19	395148	Tampa, FL	Region_19									
20	394358	Baltimore, MD	Region_20									
21	394530	Denver, CO	Region_21									
22	394982	Pittsburgh, PA	Region_22									
23	394998	Portland, OR	Region_23									
24	394458	Charlotte, NC	Region_24									
25	395045	Sacramento, CA	Region_25									
26	395055	San Antonio, TX	Region_26									
27	394943	Orlando, FL	Region_27									
28	394466	Cincinnati, OH	Region_28									
29	394475	Cleveland, OH	Region_29									
30	394735	Kansas City, MO	Region_30									
31	394775	Las Vegas, NV	Region_31									
32	394492	Columbus, OH	Region_32									
33	394705	Indianapolis, IN	Region_33									
34	395059	San Jose, CA	Region_34									
35	394355	Austin, TX	Region_35									
36	395194	Virginia Beach, VA	Region_36									
37	394902	Nashville, TN	Region_37									
38	395005	Providence, RI	Region_38									
39	394862	Milwaukee, WI	Region_39									
40	394714	Jacksonville, FL	Region_40									
41	394849	Memphis, TN	Region_41									
42	394935	Oklahoma City, OK	Region_42									
43	394807	Louisville-Jefferson County, KY	Region_43									
44	394669	Hartford, CT	Region_44									
45	394910	New Orleans, LA	Region_45									
46	395012	Raleigh, NC	Region_46									
47	394388	Birmingham, AL	Region_47									
48	394640	Grand Rapids, MI	Region_48									
49	395167	Tucson, AZ	Region_49									
50	753924	Urban Honolulu, HI	Region_50									
51	395169	Tulsa, OK	Region_51									
52	394619	Fresno, CA	Region_52									
53	395238	Worcester, MA	Region_53									
54	394415	Stamford, CT	Region_54									
55	394312	Albuquerque, NM	Region_55									
56	394308	Albany, NY	Region_56									
57	394938	Omaha, NE	Region_57									
58	394908	New Haven, CT	Region_58									
59	394357	Bakersfield, CA	Region_59									
60	394753	Knoxville, TN	Region_60									
61	394653	Greenville, SC	Region_61									
62	394952	Ventura, CA	Region_62									
63	394318	Allentown, PA	Region_63									
64	394561	El Paso, TX	Region_64									
65	394367	Baton Rouge, LA	Region_65									
66	394521	Dayton, OH	Region_66									
67	394648	Greensboro, NC	Region_67									
68	394304	Akron, OH	Region_68									
69	753906	North Port-Sarasota-Bradenton, FL	Region_69									
70	394798	Little Rock, AR	Region_70									
71	395134	Stockton, CA	Region_71									
72	394457	Charleston, SC	Region_72									
73	394484	Colorado Springs, CO	Region_73									
74	395235	Winston-Salem, NC	Region_74									
75	394440	Fort Myers, FL	Region_75									
76	394399	Boise City, ID	Region_76									
77	395160	Toledo, OH	Region_77									
78	394816	Madison, WI	Region_78									
79	394766	Lakeland, FL	Region_79									
80	394931	Ogden, UT	Region_80									
81	394528	Daytona Beach, FL	Region_81									
82	394531	Des Moines, IA	Region_82									
83	394711	Jackson, MS	Region_83									
84	394352	Augusta, GA	Region_84									
85	394666	Harrisburg, PA	Region_85									
86	395006	Provo, UT	Region_86									
87	394602	Fort Collins, CO	Region_87									

Figure 4: The multi-period portfolio selection model output for $z_{i,t}$, representing when properties are owned.

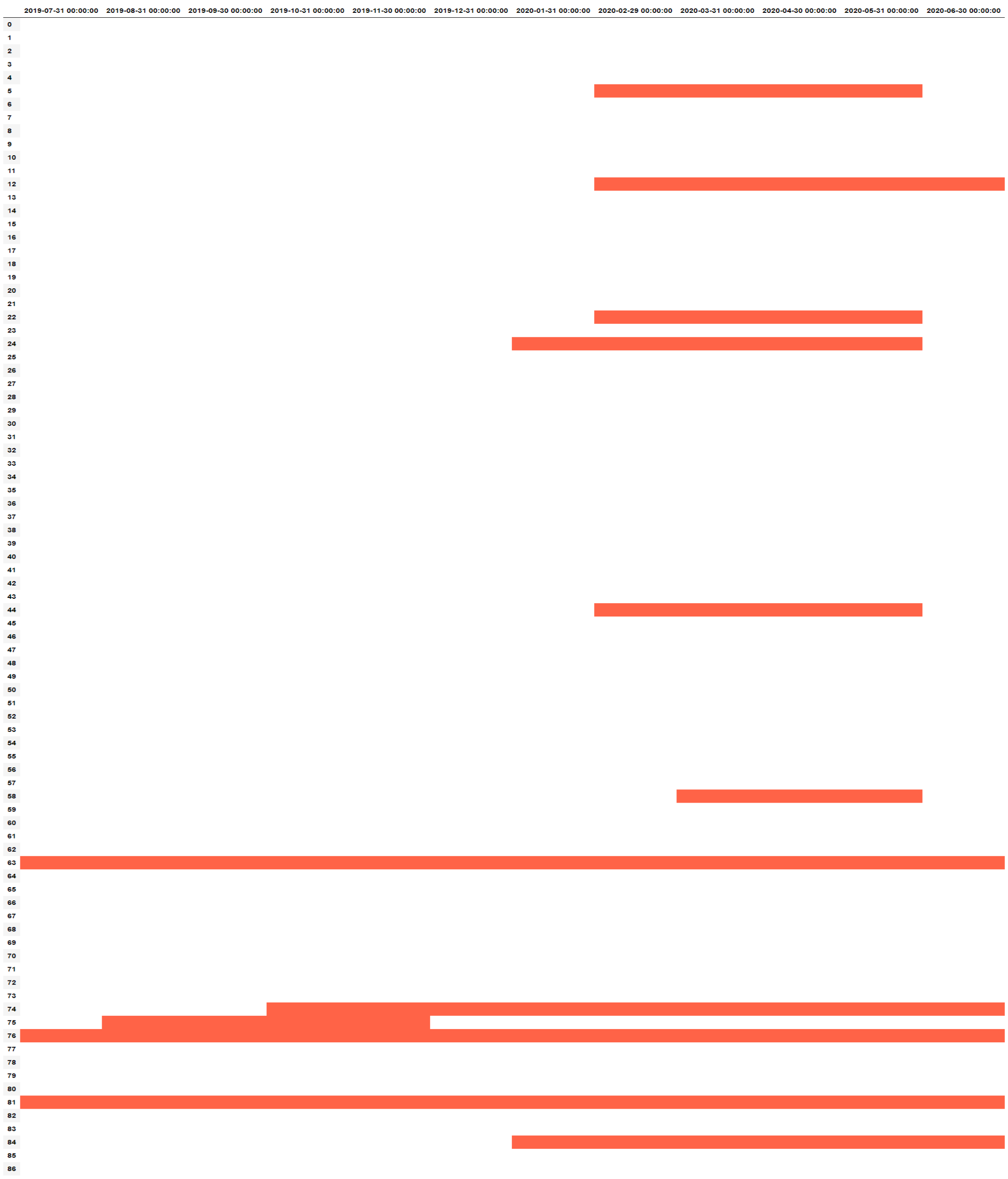


Figure 5: The multi-period portfolio selection model output for $z_{i,t}$, representing when properties are owned (continued).

Figure 6: The multi-period portfolio selection model output for $z_{i,t}$, representing when properties are owned (continued).