# Non-homogeneous Random Walks for Housing Prices

Author: Núria Foguet Coll
*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.*

Advisor: Josep Perelló Palou
(Dated: June 18, 2020)

**Abstract:** The random walk formalism has applications in many areas, with economy being one of the most notable examples. In this work, we apply this formalism to a relevant problem in modern society, such as the access to housing and its prices. A 1-D non-homogeneous random walk model serves to interpret the collected data from a citizen science public experiment on which the participants had to rent a house. Half of these games had rules which simulated regulations limiting the maximum and minimum rent prices, in order to gauge the possible effects of these policies. Although regulations do not significantly affect the prices paid by the participants, they foment demand at higher prices and reduced price variability as shown by the root mean square displacement. Finally, a bubble phase and a declining phase are distinguished, thanks to which the model of the random walk is extrapolated to larger periods of time beyond the scope of the game.

## I. INTRODUCTION

Random walks (RWs) are stochastic processes which describe the random trajectory of a particle (the walker) along a succession of steps at discrete intervals of time [1]. Historically, the study of classical one-dimensional RWs dates back as far as the 17th century, when Pierre de Fermat and Blaise Pascal worked on the 'Gambler's Ruin Problem'[1]. The mathematical theory behind RWs started to develop by Louis Bachelier's thesis on stock prices model [2]. As for the many-dimensional RWs, there were many studies around the same time. Notable examples include Lord Rayleigh's theory of sounds [4], Karl Pearson's theory of random migration of species [5], and Albert Einstein's theory of Brownian motion [6].

As said, RWs provide the framework to model a wide range of phenomena, with applications on areas as diverse as chemistry, biology, psychology or economy, to name but a few [1]. In particular, non-homogeneous RWs, where the transition probability depends on time and/or space, will be studied on the application of a one-dimensional random walk model on housing prices based on the data collected from a public participatory experiment. The data was collected by the research group OpenSystems from their public experiment 'Jocs per l'Habitatge'[8]. The experiment was an example of citizen science, and focused on the problem of housing, an issue affecting important sectors of the population. It was conducted in the three different public squares at Fort Pienc (Barcelona), Granollers and Olesa de Montserrat and consisted on a game which will be referred to as the Housing Game.

The Housing Game is a game where its 6 participants must rent a house within 12 rounds (each round can be equivalent to one month). The initial price was the official price of a 70 $m^2$ house for the location (note that in the following data analysis all the prices were normalized with respect to the initial price, so that the data from the three locations could be combined). Each round each participant must decide whether they want to rent a house or not. If none want to rent a house, the price decreases a 5%. However, if one or more players wish to rent, the fastest player to decide obtains the house and the price increases by 5%. Once a participant rents a house, they are out of the game and the game has one participant less.

Furthermore, we want to test and analyse the effect of the planned regulations by the Catalan Parliament. To that end, half of the sessions have extra rules to simulate the effect of regulations on the rent price. On the one hand, an upper threshold is established to prevent an excessive inflation of the price. On the other hand, a lower threshold is placed. The idea is to stimulate demand by conceding financial aid to those wishing to rent, such aid coming from the contribution of the other players as taxes. If either threshold were to be reached in the game, then the price would remain the same in the next round.

We here aim to study the several aspects pertaining to the RW associated to the game, such as the transition probability, the probability distribution or the root mean square displacement, in order to comprehend the evolution of the price and the demand, and the effect of regulations. Finally, after simplifying the model, the asymptotic behaviours with inflation (large demand) and deflation (low demand) are studied.

## II. RESULTS

### A. Non-homogeneous RWs

Notation wise, the rounds will be referred to as $n$, where $n \in \{0, ..., 12\}$, where $n = 0$ would be the starting point and $n = 12$ would represent the situation after the final round. Moreover, an integer $k$ will be used to refer to the price level, considering $k = 0$ initially and that each time the price rises, $k$ increases by a unit (and similarly, $k$ decreases by a unit when the price decreases). In other words, $n$ represents the number of steps taken by the walker, while $k$ is a representation of the position. Additionally, the steps are of a fixed length of one unit (when the price increases or decreases) or zero (which can only when regulations are active and one of the thresholds is reached). The probability that rules the movement of the walker each round is the transition probability or propagator $\pi_n(k'|k)$, which is the probability that at the round $n$ the price goes from $k$ to $k'$, with $k'$ being either $k$ or $k \pm 1$. It is a conditional probability computed as:

$$\pi_n(k'|k) = \frac{\#\text{ cases at }(n+1,k')\text{ from }(n,k)}{\#\text{ cases at }(n,k)} \qquad (1)$$

Due to the dependence on $n$ and $k$ of the transition probabilities, the RW is non-homogeneous. The participants being different each game, the presence of regulations and the diminishing number of participants as the houses are rented are factors contributing to said non-homogeneity.

For instance, when considering the games without regulations, $n$ and $k$ must have the same parity. Another condition that can be easily found is that $-n \leq k \leq n$. However, due to the restricted number of participants $N$, a stronger condition on the upper bound can be found. Since each time the price increases there is one less participant, necessarily $k \leq 2N - n$. Consequently, the boundary conditions are actually $-n \leq k \leq \min(n, 2N - n)$. Note that the straight line $k = 2N - n$ acts as an absorbing wall which terminates all the games reaching it, since the $N$ players will each have rented a house.

When considering regulations, the boundary conditions are given by the thresholds. The problem is that the thresholds are not univocally determined by $n$, since they depend on the sum of the times the price increases and decreases. However, it could be found that the possible values of $k$ were always among $|k| \leq 3$.

With the transition probabilities and the boundary conditions, the RW of the process can be represented. It consists of a tree of the possible $(n,k)$ points which are connected according to the different transitions, colored based on the probability of said transitions. The results without and with regulations are shown in Figures 1 and 2, respectively. The most probable paths are also represented, showing a distinct behaviour, a notable difference being the fact the with regulations the most probable path ends a round earlier. This result is consistent with the average ending rounds of the games: $\langle n \rangle = 9.9 \pm 0.4$ (with regulations) and $\langle n \rangle = 11.2 \pm 0.2$ (without regulations).

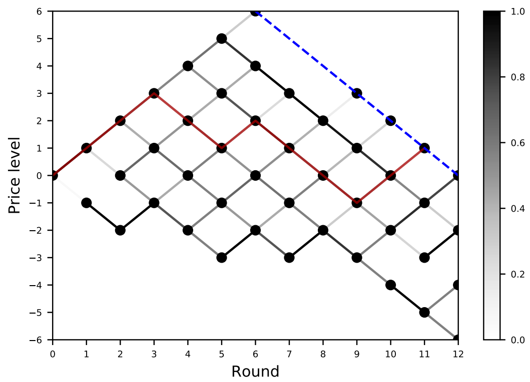


FIG. 1: RW of the process without regulations. The lines joining two points are colored according to the value of the transition probability as shown by the colorbar. Moreover, the most probable path has been highlighted in red; while the dashed blue line represents the boundary of the process.

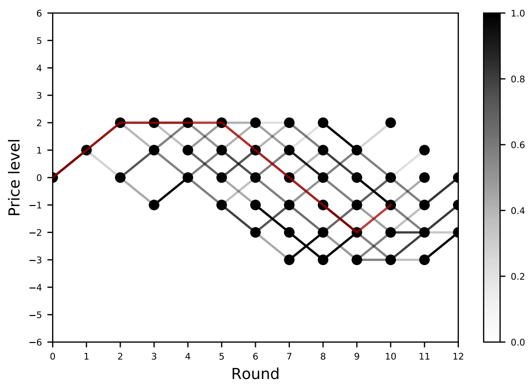


FIG. 2: RW of the process with regulations. Although for higher values of $n$ the upper threshold could technically be $k = 3$, said value is never reached.

### B. Root mean square displacement

A relevant parameter of the RW is the root mean square displacement (RMSD)

$$\sqrt{E[k(n)^2]}\,, \qquad (2)$$

which provides an idea of the distance to the starting point. A known result is that the RMSD is proportional to $\sqrt{n}$ when the transition probabilities are independent on $n$ and are both equal to $1/2$ (the RW is unbiased). Thus, by representing the RMSD in a log-log we can study the trend. The results are shown in Figure 3. For $n \leq 2$, the RMSD increases steadily with a slope of 0.89, which is greater than $1/2$ and lower than 1, implying that it is a superdiffusive process. For greater values of $n$ the dispersion is limited, specially in the case with regulations, where there is a sudden drop of the RMSD when compared with the values without regulations, implying a reduced variability in the price levels. Without regulations, the slope between $n = 2$ and $n = 5$ is lower than $1/2$, meaning it is a subdiffusive process, while for the subsequent values of $n$ barring the final ones, the slope is negative. This suggests the distinction of two phases: a bubble phase ($n \leq 2$) and a declining phase ($n \geq 3$). This division will be reinforced in the following section when studying the demand.

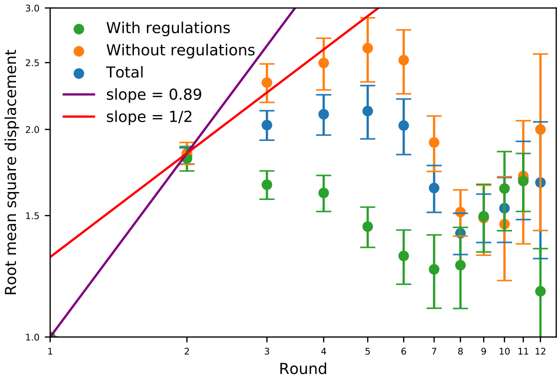


FIG. 3: Root mean square displacement for each $n$, with the standard error of the mean.

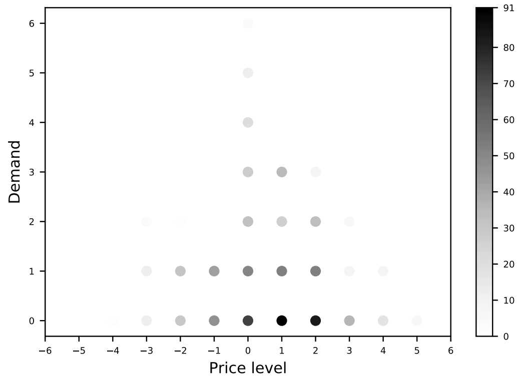


FIG. 4: Scatter plot of the demand vs the price level $k$ considering all the cases, with the color of the points related to their absolute frequency as shown by the colorbar.

### C. Demand

In the context of the game, the demand can be defined as the number of participants that each round wish to rent a house. Furthermore, as the demand reflects on the decisions of all the participants, it is an adequate way to look into the effect of socio-demographic aspects such as the gender and the age of the participants.

When studying the general behaviour of the demand, the cases with regulations, without regulations and the total cases were considered as well. Figure 4 shows the distribution of the demand with the price level considering the total cases. There is a bias towards $k > 0$, and the highest demands are reached in $k = 0$, which is reasonable due to being the initial level. It was found that with regulations the demand at larger price levels was higher than without regulations. For instance, the average demand at $k > 0$ without regulations is $0.64 \pm 0.07$ while with regulations is $0.88 \pm 0.08$. Thus, it would suggest that the regulations contribute at increasing the demand for larger prices.

The relationship between demand and round is shown by Figure 5 for the total cases. For $n \leq 2$ the demand is greater than 1, which implies that the price will increase. Thus, it reinforces the idea of a bubble phase. Moreover, for $n \geq 3$, the demand is consistently lower than 1, which suggests a declining phase when the games also have less and less participants. Note that within the final rounds in the second phase, there is a slight increase of demand due to the game nearing its end and the few remaining players still needing to rent a house. Furthermore, the average demand per round can be fitted to a polynomial of degree 4 by the least squares regression. The same behaviour can be observed when considering the cases with and without regulations separately, the only difference being that for most of the second phase the demand is slightly

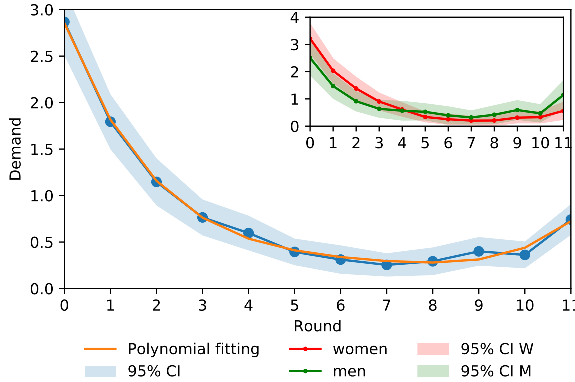


FIG. 5: Average demand versus the round, with a 95% confidence interval. In orange the polynomial fitting of the data points is represented. The same representation taking into consideration the gender of the participants is shown as an inset.

higher with regulations.

On the socio-demographic aspects of demand, the study was done with the total cases, without distinguishing whether there were regulations or not. Further dividing the samples would have resulted in some cases not having enough data. To account for the differences in sample sizes, the demand was normalized accordingly. The results in the case of gender are shown in Figure 5 as an inset. For the first few rounds, women's demand is consistently higher, while in the final rounds the trend is reversed. Regarding the differences among the several age groups, the most remarkable difference is in the first rounds where the demand for the youngest age segment (18-24 years) is the highest by a margin: at $n = 0$, the normalized demand for the 18-24 years old is $7.5 \pm 1.3$, while the second highest normalized demand at the same $n$ is a comparatively much lower $4.6 \pm 0.7$ for the segment of 25-34 years.

### D. Prices

Another aspect to consider is the probability distribution of $k$ for each $n$, $p(n,k)$. Bearing in mind the aforementioned boundary conditions, it can be calculated by:

$$p(n,k) = \frac{\#\text{ cases at }(n,k)}{\#\text{ total cases at }n} \tag{3}$$

From the propagator, $p(n,k)$ can also be calculated by means of the Master Equation:

$$p(n,k) = \sum_{k'=k,k\pm1} \pi_{n-1}(k|k')p(n-1,k') \tag{4}$$

Considering the initial condition $p(0,k) = \delta_{k=0}$, the equation can be solved recursively. Although it is simpler to use (3), the Master Equation will be useful in later sections when the continuous limit is considered.

As for the actual normalized prices paid by the participants, it was found that there were no significant differences whether there were regulations or not. The 95% confidence interval for the mean price without regulations is $1.024 \pm 0.012$, while with regulations is $1.020 \pm 0.010$, with the difference not being significant.

There were more significant differences when comparing the rent prices by gender and age group. The results for the different age groups are shown in Figure 6. The two oldest age segments have the smallest range of prices, while the position of the median shows that data of the first two segments is distributed more symmetrically. The same representation considering the gender of the participants shows that the range of rent prices of the women is wider, and the mean price they payed is higher as well when compared with men's rent price ($1.033 \pm 0.010$ versus $1.005 \pm 0.012$).

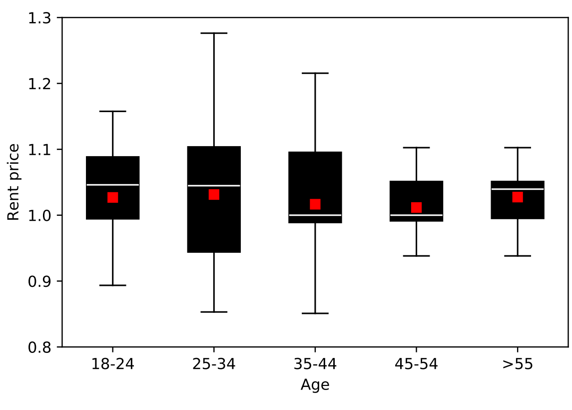


FIG. 6: Boxplot of the rent price by age. The white lines are the median while the red squares are the mean values.

## III. EXTRAPOLATING THE MODEL

As stated before when studying the RMSD and the demand, two phases can be considered. The first phase is for $n \leq 2$, where the price level increases steadily as a superdiffusive process. The second phase is for $n \geq 3$, where the price level is much more likely to decrease. Consequently, the first phase can be considered as a bubble phase with high demand, with a bubble burst in the transition from $n = 2$ to $n = 3$, and the entire second phase being a declining phase with low demand. The objective is to extrapolate both regimes in order to compare them under the same conditions. Thus, the transition probabilities of both phases are estimated. The probability that the price increases will be referred to as $p$, while $q = 1 - p$ is the probability that the price decreases. Another simplification considered is that there are no boundary conditions due to the number of participants $N$.

When considering games without regulations, the bubble phase can be approximated as a binomial distribution, while the declining phase can be approximated as the superposition of three binomial distributions due to the initial conditions at $n = 2$. However, when considering the limit of $n$ tending to infinity by means of Stirling's approximation, the initial conditions are irrelevant and the result obtained is a normal distribution

$$p(n,k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(k-n\mu)^2}{2n\sigma^2}, \tag{5}$$

where $\mu = p - q$ and $\sigma^2 = pq$, being the mean and the variance respectively for a single step of the RW.

Assuming that the first phase were to be prolonged to large values of $n$, the same distribution as in Eq. (5) is obtained, the difference being the values of $p$ and $q$. The results for a given $n$ are shown in Figure 7. As is to be expected, the bubble phase will dominate for higher values of $k$. Moreover, the dispersion is greater in the declining phase.

An alternative is to solve the Master Equation (4) in the continuous limit. To that end, one must perform the change of variables $x = k\delta x$, $t = n\delta t$, and then expanding (4) resulting in a convection-diffusion equation:

$$\frac{\partial p(t,x)}{\partial t} + v\frac{\partial p(t,x)}{\partial x} = D\frac{\partial^2 p(t,x)}{\partial x^2}, \tag{6}$$

where $D = \frac{1}{2\delta t}(\delta x)^2$ is the diffusion constant and $v = \frac{(p-q)}{\delta t}\delta x$ is the drift velocity.

With regulations, the first phase behaves similarly as without regulations, since for the first few rounds the thresholds cannot be reached. Therefore, the point of interest is the second phase in order to study the asymptotic effects of the regulations. There are constrained RW models with absorbing walls and reflective walls [3]. However, the thresholds do not fall into either category, they act as physical walls that constrain the movement of

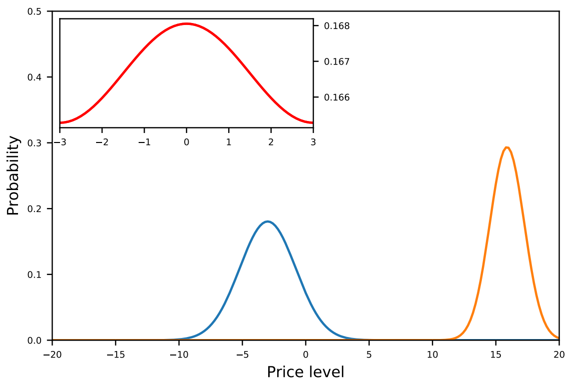


FIG. 7: Probability distribution for $n = 20$ without regulations for the 1st phase (orange) and the 2nd (blue). The inset shows the distribution for the second phase with regulations in the continuous limit for $Dt/\delta x^2 = 5$.

the walker, but do not terminate the walks upon reaching them and, unlike reflective walls, the walker may remain in the same position. A solution would be to consider reflective walls in the continuum limit and solving Eq. (6).

To further simplify the model, since $p = 0.49 \approx q = 0.51$ for the second phase, the term with $v$ from equation (6) can be ignored, yielding only a diffusion equation. By taking the boundaries at $k = \pm 3$, in the continuum $x \in [-3\delta x, 3\delta x]$. The reflective walls imply that the probability flux at the boundaries is zero. Therefore, the following solution with the initial condition $p(0, x) = \delta(x)$ can be found in the literature [9]:

$$p(t, x) = \frac{1}{6\delta x} + \frac{1}{3\delta x} \sum_{m=1}^{\infty} \cos \frac{\pi m x}{3\delta x} \exp \frac{-\pi^2 m^2 Dt}{9\delta x^2} \quad (7)$$

## IV. CONCLUSIONS

The research has explored the RW of the Housing Game, which is non-homogeneous due to participants being different and the boundary conditions because of the regulations and the finite number of participants.

The study of the RMSD has shown the relevant effect of regulations in diminishing the variability of the price by limiting the range of possible prices. Moreover, the RMSD suggested the division in a bubble and a declining phase.

Regarding the demand, its dependence on the rounds reinforced the aforementioned division of two regimes. Furthermore, the demand is biased towards positive price levels, such effect being more prominent with regulations. The participants of the game were aware of the regulations and therefore, knowing that the price would not increase beyond a certain point, they felt more inclined to rent a house. Consequently, the games with regulations finished earlier as well, thus increasing market efficiency.

However, contrary to what would be expected, the presence of regulations does not affect significantly the rent price the participants paid. The differences in rent price are in fact more significant when considering socio-demographic aspects such as the gender and the age segment of the participants. For example, women paid more than men, a result consistent with their demand being higher in the first rounds during the bubble phase.

Finally, after some simplifications, the model of the RW has been extrapolated. For both phases the steps of the RW are considered independent and identically distributed random variables, and the results without regulations are the expected from the Central Limit Theorem. Without regulations, in the high demand regime prices are higher and have lower volatility when compared to the low demand regime. With regulations the main goal is to gauge the long-term effects of the thresholds, and therefore the second phase is the focus. Due to the boundary conditions by the regulations, solving the Master Equation at the continuous limit was the best approach.

**Acknowledgments**

I would like to thank my advisor Josep Perelló for his continued support even through these trying times, as well as my family.

[1] C.H. Lo, *On some random walk problems* (Durham University, 2017).
[2] L. Bachelier, *Théorie de la spéculation* (Annales Scientifiques de l'École Normale Supérieure, 1900)
[3] J. Rudnick, G. Gaspari, *Elements of the Random Walk* (Cambridge University Press, New York 2004).
[4] J. Rayleigh, *The Theory of Sound* (Macmillan, London, 1894).
[5] K. Pearson, *A mathematical theory of random migration* (Dulau and Co., London, 1906).
[6] A. Einstein, "On the Movement of Small Particles Suspended in Stationary Liquids Required by the Molecular-Kinetic Theory of Heat". *Annalen der Physic.* **322**(8): 549-560 (1905).
[7] M. Menshikov, S. Popov and A. Wade, *Non-homogeneous random walks* (Cambridge University Press, 2016).
[8] J. Vicens, A. Cigarini and J. Perelló, *Jocs per l'Habitatge - Ciència i Acció Ciutadana* (2019). Zenodo. http://doi.org/10.5281/zenodo.3354481 [Retrieved 18th of June 2020]
[9] M. Deserno, *One-dimensional diffusion on a finite region* (Department of Chemistry and Biochemistry, UCLA, 2004)