

INSTITUTE
OF ECONOMICS



Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM

WORKING PAPER SERIES

Wealth and Price Distribution by Diffusive Approximation in a Repeated Prediction Market

Giulio Bottazzi [°]
Daniele Giachini [°]

[°] Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy

2016/13

December 2016

ISSN(ONLINE) 2284-0400

Wealth and Price Distribution by Diffusive Approximation in a Repeated Prediction Market*

Giulio Bottazzi[†] and Daniele Giachini[‡]

Istituto di Economia, Scuola Superiore Sant'Anna
Piazza Martiri della Libertà 33, 56127 Pisa - Italy

December 28, 2016

Abstract

The approximate agents' wealth and price invariant densities of a repeated prediction market model is derived using the Fokker-Planck equation of the associated continuous-time jump process. We show that the approximation obtained from the evolution of log-wealth difference can be reliably exploited to compute all the quantities of interest in all the acceptable parameter space. When the risk aversion of the trader is high enough, we are able to derive an explicit closed-form solution for the price distribution which is asymptotically correct.

Keywords: Prediction market; Repeated betting; Diffusive approximation; Fokker-Planck equation

1 Introduction

Prediction Markets, like the Iowa Electronic Market (IEM) in which people can bet on events like Presidential Elections or Congressional Control, epitomize the role of economic interaction to aggregate distributed information. Several famous economists (Arrow et al., 2008) supported the development of prediction markets

*We gratefully acknowledge the funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 640772 - DOLFINS. We thank two anonymous referees for their insightful comments and useful advice.

[†]bottazzi@sssup.it

[‡]d.giachini@sssup.it

based on their supposed ability to provide accurate evaluations of the likelihood of uncertain events. Recently, the attention has shifted to the case of repeated prediction markets, which share more similarities with other financial market models. Beygelzimer et al. (2012) and Kets et al. (2014) study a simple repeated prediction market in which bets on binary outcomes are traded in discrete time. In particular, they analyze the case in which two agents with different beliefs about the probability of the outcomes, split their investment between the risky bet and a risk free security, according to the fractional Kelly rule.

The authors derive their most relevant conclusions about agents' wealth and bet price performing extensive simulations. Here we propose a different approach: instead of computing the quantities of interests averaging over long enough artificial time series, we suggest to compute them from the invariant distribution derived from the continuous time diffusive approximation of the discrete process.

The diffusive approximation is obtained postulating that a Poisson arrival drives the successive steps of the discrete time process, hence deriving the master equation in continuous time and finally truncating the associated Kramers-Moyal expansion at the second order.

The advantages of this procedure are many. First, the invariant distribution can be used to analyze the qualitative properties of the model, in particular whether, conditional on their beliefs and risk aversion, the two agents both stay in the market indefinitely or, conversely, if one of them will end up having all the wealth. Second, the obtained approximation fits very well the invariant distribution of the true discrete process, and can be used to compute with high precision all the statistics of interest. In addition, when the risk aversion of agents is sufficiently high, one is also able to derive the analytic expression for the approximate invariant price distribution. Third, in the limit in which the risk aversion of agents goes to infinity, the diffusive approximation becomes asymptotically correct and one can easily prove the conjecture, advanced by Kets et al. (2014), that the price converge in distribution to the true probability of the process.

Beyond this simple application, we believe that the procedure proposed in the present paper can be effectively applied to the qualitative and quantitative analysis of similar heterogeneous agents models.

2 The model

Time is discrete and at each time step t a certain event can occur with probability π^* . We set $s_t = 1$ if at time t the event occurs and $s_t = 0$ if it does not.¹ Consider two agents who repeatedly exchanges bets on this binary event. They can choose

¹This simple stochastic process is usually named "Bernoulli trials".

to gamble on the occurrence of the event or against it. The amount of wealth which is not wagered is considered invested in a riskless security, on which, without loss of generality, we assume no interest is paid. The total amount bet is redistributed among the agents proportionally to how much they have wagered, that is according to the procedure commonly known as *parimutuel*. At each time step, the agents bet on the outcome of the process by exchanging two short-lived assets in unit supply: asset 1 pays 1 only when $s_t = 1$ and asset 2 pays 1 only when $s_t = 0$. This means that if agent i invest a fraction α_t^i of her wealth on the first asset and a fraction $1 - \alpha_t^i$ on the second, she is actually investing a fraction $1 - \alpha_t^i$ on the riskless security and, at the same time, she is taking a position $2\alpha_t^i - 1$ on the occurrence of the event: if $\alpha_t^i > 1/2$ she is betting on its realization, while if $\alpha_t^i < 1/2$ she is betting against it.²

We further assume that the two agents do not know the true value π^* but rather possess different individual opinions about the probability of the event, π^i , with $i \in \{1, 2\}$. Moreover, they behave according to a so-called fractionally Kelly rule (MacLean et al., 1992, 2005), investing a fraction of wealth

$$\alpha_t^i = \alpha^i(p_t) = c\pi^i + (1 - c)p_t, \quad (1)$$

on the first asset and what remains, $1 - \alpha^i(p_t)$, on the second. The rule (1) correspond to a linear combination of the Kelly rule, shortly described by the prescription “bet your believes”, and the risk-less strategy of investing proportionally to the prevailing market price. The mixing coefficient c is assumed equal among the two bettors, and without loss of generality we pose $\pi^2 > \pi^1$. Since the lower the value of c , the higher the proportion of wealth invested in the riskless security, this parameter can be considered a measure of agents’ risk aversion. The system is closed since no wealth is consumed and all the wealth wagered is redistributed. Normalizing the total wealth to 1, denoting with w_t the wealth of the first agent and calling p_t the price of asset 1, the amount of asset 1 purchased by agent 1 can be written $\alpha^1(p_t)w_t/p_t$ while the amount purchased by agent 2 is $\alpha^2(p_t)(1 - w_t)/p_t$. Setting the asset’s supply equal to its demand one has

$$p_t = \pi^1 w_{t-1} + \pi^2 (1 - w_{t-1}). \quad (2)$$

An analogous argument and a simple computation show that the price of asset 2 is $1 - p_t$. At the end of each period the outcome of the event is revealed and assets pay out: an agent gets a fraction of the unitary payoff equal to the quantity of the

²Bottazzi and Giachini (2016) show that the model discussed here is equivalent, through a simple change of variables, to the repeated prediction market model discussed in Kets et al. (2014).

asset purchased. Thus the wealth of agent 1 evolves according to

$$w_t = \begin{cases} \frac{\alpha_t^1 w_{t-1}}{p_t} & \text{if } s_t = 1, \\ \frac{(1 - \alpha_t^1) w_{t-1}}{1 - p_t} & \text{if } s_t = 0. \end{cases} \quad (3)$$

3 Poisson arrival and relative wealth distribution

Dealing with discrete time and continuous-state space Markov chain is generally difficult. The reason is that the chain only explores a countable subset of states and the concepts of irreducible sets and ergodic measure have to be generalized (Tweedie and Meyn, 2009). For this reason, the identification of a continuous time process that approximates the discrete one can hugely simplify the analysis. We perform this identification by introducing a simple continuous-time random walk (Scalas, 2006) version of the discrete process and then considering its diffusive limit. For this purpose, it is convenient to rewrite the evolution of the model in terms of the relative log wealth $z_t = \log w_t - \log(1 - w_t)$ so as to obtain an unbounded process on the real line. This model is discrete in time and we proceed to obtain a continuous time version using a homogeneous Poisson process.³ Assume that in a short period of time δt there is a probability $\lambda \delta t$ that a new trading round takes place, a new state of the world is realized and the system is updated according to $z_{t+\delta t} = g(z_t, s_t)$ with

$$g(z, s) = z + \log \frac{\alpha^1(p)(2s - 1) + 1 - s}{\alpha^2(p)(2s - 1) + 1 - s}, \quad \text{and} \quad p = \frac{\pi^1 e^z + \pi^2}{1 + e^z}.$$

At the same time, there is a probability $1 - \lambda \delta t$ that nothing happens and $z_{t+\delta t} = z_t$. In terms of the probability density of the process $f_z(x, t) = d\text{Prob}\{z_t \leq x\}/dx$, the infinitesimal Chapman-Kolmogorov equation reads

$$f_z(x, t + \delta t) = (1 - \lambda \delta t) f_z(x, t) + \lambda \delta t \int dy \int d\pi(s) f_z(y, t) \delta(x - g(y, s)),$$

where $\delta(\cdot)$ is the Dirac delta function and $\pi(s)$ the probability measure of the underlying state process. Re-arranging terms and taking the limit for $\delta t \rightarrow 0$ one obtains the master equation

$$\frac{\partial}{\partial t} f_z(x, t) = \lambda \int dy f_z(y, t) \mathcal{K}(x, y), \quad \mathcal{K}(x, y) = \int d\pi(s) \delta(x - g(y, s)) - \delta(x - y).$$

³This ‘‘Poissonization’’ trick has been used in the early analysis of the asymptotic behavior of goodness of fit statistics, see Kolmogorov (1933) and Kac (1949)

By truncating its Kramers-Moyal expansion

$$\frac{\partial}{\partial t} f_z(x, t) = \lambda \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial x} \right)^n D_n(x) f_z(x, t), \quad D_n(x) = \frac{1}{n!} \int dy (y - x)^n \mathcal{K}(y, x)$$

at the second term, the Fokker-Planck equation is derived,

$$\frac{\partial}{\partial t} f_z(x, t) = -\frac{\partial}{\partial x} \left(\lambda D_1(x) f(x, t) - \frac{\partial}{\partial x} (\lambda D_2(x) f(x, t)) \right), \quad (4)$$

with

$$D_1 = \left[\pi^* \log \frac{c\pi^1 + (1-c)p(x)}{c\pi^2 + (1-c)p(x)} + (1 - \pi^*) \log \frac{1 - c\pi^1 - (1-c)p(x)}{1 - c\pi^2 - (1-c)p(x)} \right] \quad \text{and}$$

$$D_2 = \frac{1}{2} \left[\pi^* \left(\log \frac{c\pi^1 + (1-c)p(x)}{c\pi^2 + (1-c)p(x)} \right)^2 + (1 - \pi^*) \left(\log \frac{1 - c\pi^1 - (1-c)p(x)}{1 - c\pi^2 - (1-c)p(x)} \right)^2 \right].$$

4 Persistent heterogeneity

Since the process is unbounded, we can assume the natural boundary conditions⁴ so that if the process in (4) admits an invariant distribution, the related density should read

$$f_z(x) = \frac{f_0}{D_2(x)} \exp \left(\int_{x_0}^x dy \frac{D_1(y)}{D_2(y)} \right), \quad (5)$$

where f_0 is a normalization constant. The existence of an invariant distribution implies the possibility to attain a statistical equilibrium condition in which the unconditional expectation of the net flow of wealth between the two agents is zero (Garibaldi and Scalas, 2010). The expression in (5) constitutes an acceptable probability density if and only if

$$\lim_{x \rightarrow \pm\infty} D_1(x) = \lim_{x \rightarrow \pm\infty} \mathbb{E}[z_{t+1} - z_t | z_t] \leq 0. \quad (6)$$

In fact, as discussed in Bottazzi and Giachini (2016), the system has two possible long-run outcomes: or one agent ends up owning the entire wealth, that is $\lim_{t \rightarrow \infty} w_t = 0, 1$, and in this case the market price converges to the belief of that agent, $\lim_{t \rightarrow \infty} p_t = \pi^2, \pi^1$, or, alternatively, both agents stay in the market in the long run, their wealth shares persistently fluctuate and p_t keeps moving

⁴The probability current is asymptotically vanishing in all cases in which the conditions in (7) are satisfied.

in the interval (π^1, π^2) . The second outcome constitutes a situation of persistent heterogeneity, in which agents with different beliefs can indefinitely coexist in the market, and (6) is a sufficient and necessary condition for its occurrence. If instead $\lim_{x \rightarrow \pm\infty} D_1(x) > 0$, then the first agent asymptotically acquires all wealth, $\lim_{t \rightarrow \infty} w_t = 1$, while if $\lim_{x \rightarrow -\infty} D_1(x) < 0$ is the second agent to do it and $\lim_{t \rightarrow \infty} w_t = 0$. These two cases are associated to degenerate densities with mass in $+\infty$ or $-\infty$ respectively. In terms of the parameters of the model, (6) read

$$\begin{aligned} \pi^* \log(c\pi^j + (1-c)\pi^i) + (1-\pi^*) \log(1 - c\pi^j - (1-c)\pi^i) \geq \\ \pi^* \log(\pi^i) + (1-\pi^*) \log(1 - \pi^i), \end{aligned} \quad (7)$$

for $i, j = 1, 2$ and $i \neq j$. Notice that (7) is always satisfied for c sufficiently small, as long as π^1 and π^2 are one larger and one smaller than π^* . Remarkably, the condition for the persistent heterogeneity derived by imposing the existence of the invariant distribution of the diffusive approximation are identical to those derived for the original process by Bottazzi and Giachini (2016) applying the results based on martingale convergence theorem in Bottazzi and Dindo (2015). In this respect (4) perfectly replicates the qualitative behavior of the original, discrete time, model.

5 Approximate invariant density and price distribution

The integral in (5) can be easily performed numerically for any acceptable parametrization. This is much faster than performing Monte Carlo simulations⁵ and has the advantage of directly providing an approximation of the long-run invariant distribution. As can be seen in Fig. 1, the agreement of (5) with the invariant distribution of wealth is very good, for any set of parameter values considered. This agreement can be quantified using the Kolmogorov-Smirnov statistics. The analysis for $\pi^1 = 0.3$ and $\pi^2 = 0.8$ is reported in Fig. 2. The difference between (5) and the distribution obtained with extensive simulations is basically never rejected at a confidence level of 5%. This is a generic result and it applies robustly for all couples of values for π^1 and π^2 as long as (6) is satisfied.

The dynamics of the system can also be studied in terms of price. From (1), (2) and (3) it is

$$p_{t+1} = p_t + \frac{c(\pi^2 - p_t)(p_t - \pi^1)}{p_t + s_t - 1}. \quad (8)$$

⁵Due to the strong autocorrelation of the process, especially when c is small, one has to generate very long series before obtaining reliable Monte Carlo estimates.

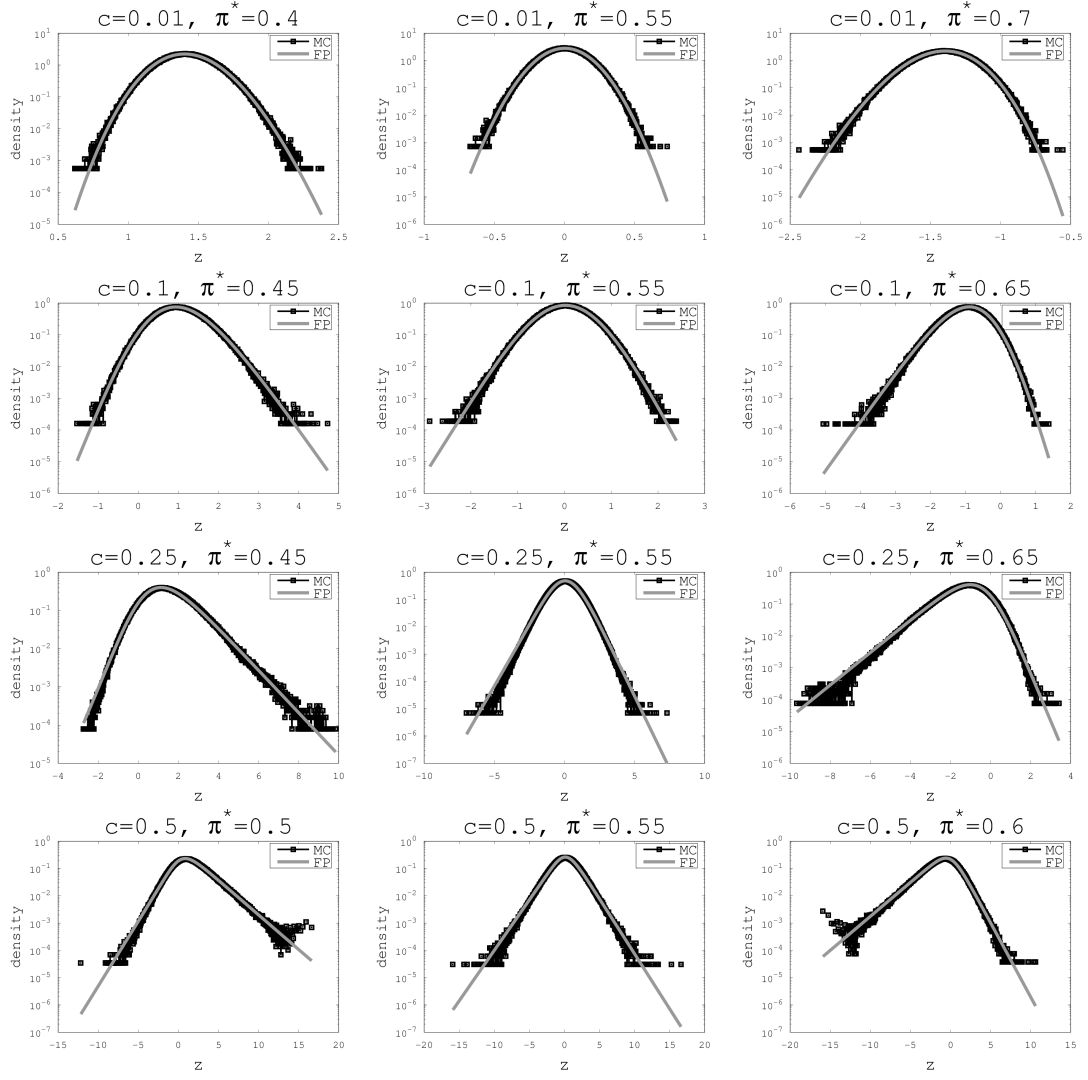


Figure 1: Approximated (FP) and simulated (MC) invariant density for $\pi^1 = 0.3$, $\pi^2 = 0.8$ and different values of c and π^* in semi-log scale. In these plots, the difference between the approximated and the simulated distributions is on average 0.0059 and never larger than 0.0561.

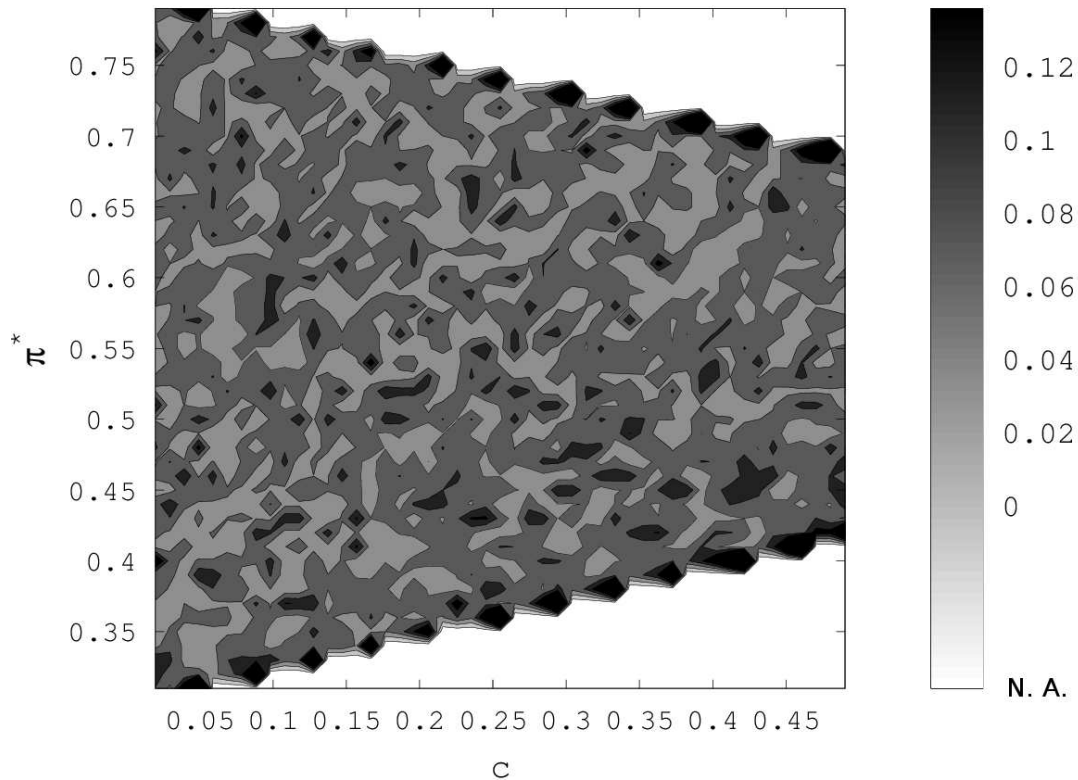


Figure 2: Kolmogorov-Smirnov D statistics between the theoretical approximation (5) and the empirical distribution obtained via extensive Monte Carlo simulations, for different values of c and π^* with $\pi^1 = 0.3$ and $\pi^2 = 0.8$. For each combination of c and π^* we considered 100 independent realizations for 10000 steps using π^* as initial condition. The maximum of the colorbar has been set to 0.1358, the Kolmogorov-Smirnov critical value for a 95% confidence level (adjusted for the sample size) for a sample of size 100. Thus, black areas can be interpreted as the cases in which the hypothesis of equality between the approximated expression and the true distribution is rejected with a confidence level of 95%. Their casual pattern suggests a very good agreement. Indeed, if the distributions were the same, we should expect a 95% confidence rejection in the 5% of cases. White areas represent set of points for which (6) is not satisfied.

Again, postulating a Poisson arrival of trading rounds we can obtain a continuous time master equation for $f_p(x, t) = d\text{Prob}\{p_t \leq x\}/dx$. The Kramers-Moyal expansion becomes now a power expansion in c

$$D_n^p(x) = c^n \int d\pi(s) \left(\frac{(\pi^2 - x)(x - \pi^1)}{x + s - 1} \right)^n = c^n M_n(x). \quad (9)$$

So we expect the diffusive approximation to improve the lower the value of c . At the same time, however, the diffusive approximation based on price dynamics breaks down for higher value of c . To see it, let us first formally derive the stationary solution of the Fokker-Planck equation (4) for f_p with the coefficients as in (9). It turns out that in this case the integral in the exponent of (5) can be computed explicitly to give

$$f_p(x) = \frac{2 f_0}{c^2} \frac{(x - \pi^1)^{2\frac{A_1}{c} - 2}}{(\pi^2 - x)^{2\frac{A_2}{c} + 2}} (\pi^*(1 - \pi^*) + (\pi^* - x)^2)^{-\frac{1+A_1-A_2}{c} - 1} x^2 (1 - x)^2 \exp \left\{ 2 \left(\frac{A_2}{\pi^2 - \pi^*} + \frac{A_1}{\pi^* - \pi^1} \right) \frac{\sqrt{\pi^*(1 - \pi^*)}}{c} \arctan \left(\frac{\pi^* - x}{\sqrt{\pi^*(1 - \pi^*)}} \right) \right\}, \quad (10)$$

where f_0 is a normalization constant and with

$$A_i = \frac{(\pi^* - \pi^i)(1 - \pi^i)\pi^i}{(\pi^2 - \pi^1)(\pi^*(1 - \pi^*) + (\pi^* - \pi^i)^2)}, \quad i = 1, 2.$$

The formal solution (10) represents a good approximation of the invariant distribution of the bounded price process only if the associated potential $\Phi_p(x) = -\log f_p(x)$ diverges to $+\infty$ at the boundaries of the interval (π^1, π^2) , that is when the reflecting barrier hypothesis is satisfied (Risken, 1984). From (10) it is immediate to see that

$$\lim_{x \rightarrow \pi^1+} \Phi_p(x) \sim \lim_{x \rightarrow \pi^1+} \left(1 - \frac{A_1}{c} \right) \log(x - \pi^1),$$

$$\lim_{x \rightarrow \pi^2-} \Phi_p(x) \sim \lim_{x \rightarrow \pi^2-} \left(1 + \frac{A_2}{c} \right) \log(\pi^2 - x).$$

Notice that A_1 is by definition positive and A_2 negative. Thus, we can conclude that (10) represents a good approximation only when $c < A_1, -A_2$. It is important to remark that the absorbing barrier condition, that is a potential diverging to minus infinity, which characterizes (10) for larger values of c , is a mere artifact of the diffusive approximation. In fact, as long as (7) are satisfied, we know that the process is stationary and both barriers are reflecting. This suggests that the diffusive

approximation of the price dynamics is less reliable than the diffusive approximation of the wealth ratio: when c is large, the truncation of the Kramers-Moyal expansion at the second term discards relevant information about the behavior of the price process and misinterprets the nature of the boundaries.

We conclude the paper addressing the hypothesis, advanced in Kets et al. (2014), that when the mixing parameter becomes asymptotically small, that is $c \rightarrow 0$, then the prevailing market price converges to the true value, $p \rightarrow \pi^*$. Formally, the invariant price density, with reference to (4) and (9), can be written as

$$f_p(x) = \frac{f_0}{c^2 M_2(x)} \exp\left(\frac{1}{c} \int_{\pi^*}^x dy \frac{M_1(y)}{M_2(y)}\right).$$

Notice that $M_1(x) > 0$ when $x < \pi^*$ and $M_1(x) < 0$ when $x > \pi^*$, hence the expression in the exponent is never positive and has its maximum for $x = \pi^*$. We can thus apply the asymptotic expansion of the Laplace-type integral to obtain

$$\lim_{c \rightarrow 0} \int dx f_p(x) x^n \sim (\pi^*)^n f_0 \sqrt{\frac{2\pi}{c^3 M_2(\pi^*) |M_1'(\pi^*)|}}$$

which, by the normalization condition of the probability density, reduces to

$$\lim_{c \rightarrow 0} \int dx f_p(x) x^n \sim (\pi^*)^n$$

implying that $\lim_{c \rightarrow 0} f_p(x) = \delta(x - \pi^*)$, confirming the hypothesis by Kets et al. (2014).

6 Conclusion

The diffusive approximation of both the wealth and the price invariant distribution for the repeated prediction market model of Kets et al. (2014) were derived. Our method relies on embedding the discrete time model in continuous time via a Poisson process and computing the diffusive approximation of the state variable of interest via the Fokker-Planck equation. The derived approximation allows for a simple discussion of the asymptotic behavior of the model. Moreover, we show that the diffusive invariant distribution of the log wealth ratio approximates very well the real one on the relevant region of the parameter space where (7) are satisfied, so that the asymptotic distribution of all other quantities can be reliably derived via a density function transformation. We also obtained a closed form expression for the diffusive approximation of the price invariant density when the mixing parameter c is sufficiently small, that is agents' risk aversion is high. This expression confirms the hypothesis, advanced in Kets et al. (2014), that the

prevailing market price converges to the true probability value when the value of c goes to zero.

References

- Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., Levmore, S., Litan, R., Milgrom, P., Nelson, F. D., et al., 2008. The promise of prediction markets. *SCIENCE-NEW YORK THEN WASHINGTON-320* (5878), 877.
- Beygelzimer, A., Langford, J., Pennock, D. M., 2012. Learning performance of prediction markets with kelly bettors. In: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*. International Foundation for Autonomous Agents and Multiagent Systems, pp. 1317–1318.
- Bottazzi, G., Dindo, P., 2015. Drift criteria for persistence of discrete stochastic processes on the line. Tech. rep., Institute of Economics, Sant’Anna School of Advanced Studies, Pisa, Italy.
- Bottazzi, G., Giachini, D., 2016. Far from the madding crowd: Collective wisdom in prediction markets. *LEM Papers Series 2016/14*, Institute of Economics, Sant’Anna School of Advanced Studies, Pisa, Italy.
- Garibaldi, U., Scalas, E., 2010. *Finitary Probabilistic Methods in Econophysics*. Cambridge University Press, Cambridge UK.
- Kac, M., 1949. On deviations between theoretical and empirical distributions. *Proceedings of the National Academy of Sciences of the United States of America* 33 (5), 252–257.
- Kets, W., Pennock, D. M., Sethi, R., Shah, N., 2014. Betting strategies, market selection, and the wisdom of crowds. In: *Twenty-Eighth AAAI Conference on Artificial Intelligence*.
- Kolmogorov, A. N., 1933. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell’Istituto Italiano degli Attuari* 4, 83–91.
- MacLean, L., Ziemba, W. T., Blazenko, G., 1992. Growth versus security in dynamic investment analysis. *Management Science* 38 (11), 1562–1585.
- MacLean, L. C., Ziemba, W. T., Li, Y., 2005. Time to wealth goals in capital accumulation. *Quantitative Finance* 5 (4), 343–355.

Risken, H., 1984. Fokker-planck equation. Springer.

Scalas, E., 2006. The application of continuous-time random walks in finance and economics. *Physica A* 362, 225–239.

Tweedie, R. L., Meyn, S., 2009. *Markov Chains and Stochastic Stability*. Cambridge University Press, Cambridge UK.