Dynamic Pricing with Model Uncertainty: Learning & Earning

(tutorial proposal at *ACM EC 2015*)

Aleksandrs Slivkins*		Assaf Zeevi†

1 Tutorial description

This tutorial focuses on *dynamic pricing under model uncertainty*: a class of problems whose first instance dates back at least 40 years, is relatively simple in structure, is widely considered fundamental, and has numerous manifestations across multiple application domains and academic disciplines. While significant progress has been made throughout the last several decades, including a flurry of recent work, many variants of this problem class remain essentially unsolved. Briefly stated, the problem can be described as sequential pricing when the underlying demand model (or demand curve) is unknown and the market response to any given price is confounded by statistical noise. It will be helpful to hold in mind the following simple problem instance. The decision maker ("seller") faces demand for a product s/he is selling. At every successive time unit the seller fixes a price for the product, subsequent to which demand is realized. The demand realizations are "noisy" observations of an ambient demand curve which is unbeknownst to the seller. The seller's objective is to maximize expected cumulative (either discounted or not) profits over the time horizon that governs the interactions with the buyers. This situation is a quintessential example of a trade-off between exploration of the environment (to learn demand characteristics) and exploitation of that knowledge (via pricing) to maximize expected rewards.

General background. In economics, probably the most influential and frequently cited work in this context is [48]; see also [26, 1]. An infinite-horizon, time-discounted Bayesian formulation was employed, and the authors focused on the following question: is it certain that a seller who follows an ex ante optimal policy will eventually obtain complete information about the underlying demand environment? This question, termed the *learnability problem*, was resolved in the negative in various settings. In [48] prices were restricted to a finite set, and the setting was formulated as a multi-armed bandit problem (MAB); further connections to the MAB literature will be expounded on later. The learnability problem has close ties with various current strands of work in machine learning, adaptive control, and dynamical systems. Examples of further work in the economics community includes [34, 35].

In operations research and management science (OR/MS), dynamic pricing is commonly studied under the umbrella of "revenue management", typically with *capacity constraints*: limited supply of goods to be sold. In this context, [7] is among the first OR/MS papers to consider model uncertainty. They consider a Bayesian model for consumer demand, with a single unknown parameter and conjugate priors, and rely on dynamic programming methods; see [28] for a more recent paper of similar character. [14] depart from the Bayesian formulations, treating both parametric model uncertainty and the case in which the demand model need not belong to any parametric family. If sufficient restrictions are made on the problem primitives, it may be possible, at least in the Bayesian formulation, to characterize the optimal policy via solution of the

*Microsoft Research, New York NY, USA. Email: slivkins@microsoft.com.
†Columbia University, New York NY, USA. Email: assaf@gsb.columbia.edu

corresponding Bellman equation. More realistically, the papers in this strand of literature focus on deriving *near-optimal* solutions; pricing policies whose performance can be guaranteed to be within some "distance" of the optimum; see [15, 16, 17, 36, 20, 24, 53, 49, 18, 27, 46, 45] for recent contributions in this spirit.

Dynamic pricing under mode uncertainty has also attracted significant attention in computer science, starting from [19, 39]. Focusing on prior-independent formulations, this line of work obtains successively more and more general results [9, 10, 12, 13] in terms of inventory constraints and other aspects. It has also considers a variant in which the decision maker is buying rather than selling [11, 50, 12, 13, 33]; this variant is particularly relevant to pricing tasks in crowdsourcing markets, see [52] for more background.[1] Further, dynamic pricing is one of the main applications for MAB problems with continuous action spaces, e.g. [2, 38, 6, 41, 40, 42, 21, 47, 51]. One major theme in all this work (both on dynamic pricing and on MAB) is *adaptive exploration*: adapting the exploration schedule over time depending on the observations, so that the apparently suboptimal choices are gradually phased out. Also, as it is often useful to "discretize" the action space – only focus on a small, finite subset thereof – another major issue is how to choose such "discretization" in a principled way. The discretization can be either fixed in advance, or *adaptive*: refined over time depending on the observations so as to zoom in on the more promising regions of the action space.

A brief map of the problem space. In the most basic formulation, a seller interacts with one customer per round, and offers her a single item (at the chosen price). This version can be extended along several "modeling dimensions", some of which we outline below. First, the unlimited supply of goods [19, 39, 27, 20] may be restricted to hard supply constraints [14, 54, 10, 16, 12], which in turn can be relaxed to concave constraints [8]. Second, rather than a single product there may be multiple products (e.g., [23, 37]), and more generally *production networks* [46, 45, 16, 12] where a limited supply of some "primitive resources" may be used to produce the goods. Third, in each round the seller may provide a single offer or a menu of options to choose from, and each offer may contain multiple items (of identical or different products) that are priced with volume- or bundle-based discounts or surcharges. Fourth, there is a wide spectrum that lies between fully known and completely unknown demand model; model uncertainty may be parametric or non-parametric. Fifth, demand may change over time, either in a completely adversarial way [39] or with significant restrictions on the per-round change [51], the number of changes [15, 49], or the total change over the time horizon [37]. Finally, each customer may come with a known *context*, such as a customer profile or purchase history, and the seller's actions may depend on this context [29, 51, 30, 13]. The last two "dimensions" are closely related to, respectively, MAB with change over time (e.g. [5, 31, 51]), and MAB with contexts (e.g. [5, 4, 32, 44, 25]).

Most of the above "modeling dimensions" carry over to the "buy rather than sell" version; here selling multiple products corresponds to commissioning multiple types of tasks.

Open questions. This problem space is quite rich in open questions; we will outline some of them.

Some of the main "research frontiers" on the above map concern limited-supply dynamic pricing. First, there is virtually no work that combines limited supply and change over time; it unclear how well performing policies might look in that context. Second, fixed discretization is poorly understood for settings with multiple products (not to mention computationally intensive) and adaptive discretization can be elusive even for a single product. Third, it is unclear how to optimally incorporate contexts into pricing policies; in this case a significant additional layer of complexity is that well-performing pricing policies may be computationally intractable.

Other key open questions have a more descriptive flavor. An illustrative example of this variety is the effect of model misspecification (true demand curve is not nested within the family of assumed demand models), both on price dynamics and on the resulting revenue.[2] A second question concerns the value of

[1] While the "selling" and "buying" versions are technically different, some models subsume both versions [12, 13].

[2] A particularly striking consequence of model misspecification, termed the "spiral down effect", has been documented in [22].

side information (contexts) and, in particular, how this value is affected by the structure of the underlying demand curve. Third, a long-standing open question concerns the celebrated conjecture of Anderson and Taylor [3] on whether a myopic pricing policy (namely, one based on iterated least squares) is asymptotically optimal. Despite a partial negative response by Lai and Robbins [43] in the context of MAB, this question has not yet been resolved in a satisfactory manner for dynamic pricing.

Structure of the proposed tutorial. We cover in detail the basic version of the dynamic pricing under model uncertainty, and use this version to show-case some of the major themes and techniques in this problem space. We outline the various "modeling dimensions" described above, and highlight several key results. Along the way, we emphasize the contributions and influences from different areas (economics, OR/MS and computer science), and hint at connections to other areas such as adaptive control, dynamical systems, and mechanism design. We conclude with some open questions.

2 Tutor biographies

Alex Slivkins is a research scientist at Microsoft Research New York. Alex's research interests are in algorithms and theoretical computer science, spanning machine learning theory, social network analysis, and algorithmic economics. He has also worked on metric embeddings and algorithms for Internet and peer-to-peer networks. Across various domains, Alex is drawn to algorithmic problems with informational constraints. He is particularly interested in sequential decision-making and its applications to web search, mechanism design, and crowdsourcing markets. His work has received the best paper award at *ACM EC 2010* and the best student paper award at *ACM PODC 2005*.

Before joining MSR New York in 2013, Alex Slivkins was a member of MSR Silicon Valley since 2007. Alex received his Ph.D. in Computer Science from Cornell University in 2006, under the advisorship of Jon Kleinberg. In 2006-2007 he was a postdoc at Brown University, supervised by Eli Upfal. Alex's undergraduate degree is from Caltech, B.S. in Mathematics with distinction.

Assaf Zeevi is Vice Dean for Research and Henry Kravis Professor at the Graduate School of Business, Columbia University. His research is broadly focused on the formulation and analysis of mathematical models of complex systems, with particular interest in the intersection of Operations Research, Statistics, Computer Science and Economics. Assaf received his B.Sc. and M.Sc. (1997) Cum Laude from the Technion, in Israel, and subsequently his Ph.D. (2001) from Stanford University. In 2001 he joined the faculty of the Business School at Columbia University. He is a recipient of a CAREER Award from the National Science Foundation, an IBM Faculty Award, Google Research Award, the INFORMS Revenue Management Society (2008) and M&SOM (2013) Best Publication Awards, and the Dean's Award for Teaching Excellence. Assaf consults with various companies in the areas of high technology, financial services, and revenue management, and serves on several scientific advisory boards. He is currently a member of the INFORMS Applied Probability Society Prize Committee, and serves on the editorial boards of several leading journals in his profession.

References

[1] P. Aghion, P. Bolton, C. Harris, and B. Jullien. Optimal Learning by Experimentation. *The Review of Economic Studies*, 58(4):621–654, 1991.

[2] Rajeev Agrawal. The continuum-armed bandit problem. *SIAM J. Control and Optimization*, 33(6):1926–1951, 1995.

[3] T. W. Anderson and J. Taylor. Some Experimental Results on the Statistical Properties of Least Squares Estimates in Control Problems. *Econometrica*, 44(6):1289–1302, 1976.

[4] Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *J. of Machine Learning Research (JMLR)*, 3:397–422, 2002. Preliminary version in *41st IEEE FOCS*, 2000.

[5] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM J. Comput.*, 32(1):48–77, 2002. Preliminary version in *36th IEEE FOCS*, 1995.

[6] Peter Auer, Ronald Ortner, and Csaba Szepesvári. Improved Rates for the Stochastic Continuum-Armed Bandit Problem. In *20th Conf. on Learning Theory (COLT)*, pages 454–468, 2007.

[7] Y. Aviv and A. Pazgal. A Partially Observed Markov Decision Process for Dynamic Pricing. *Management Science*, 51(9):1400–1416, 2005.

[8] Yossi Azar, Uriel Feige, Michal Feldman, and Moshe Tennenholtz. Sequential decision making with vector outcomes. In *Innovations in Theoretical Computer Science Conf. (ITCS)*, pages 195–206, 2014.

[9] Moshe Babaioff, Liad Blumrosen, Shaddin Dughmi, and Yaron Singer. Posting prices with unknown distributions. In *Innovations in Theoretical Computer Science Conf. (ITCS)*, 2011.

[10] Moshe Babaioff, Shaddin Dughmi, Robert D. Kleinberg, and Aleksandrs Slivkins. Dynamic pricing with limited supply. 3(1):4, 2015. Special issue for *13th ACM EC*, 2012.

[11] Ashwinkumar Badanidiyuru, Robert Kleinberg, and Yaron Singer. Learning on a budget: posted price mechanisms for online procurement. In *13th ACM Conf. on Electronic Commerce (EC)*, pages 128–145, 2012.

[12] Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. Bandits with knapsacks. In *54th IEEE Symp. on Foundations of Computer Science (FOCS)*, 2013.

[13] Ashwinkumar Badanidiyuru, John Langford, and Aleksandrs Slivkins. Resourceful contextual bandits. In *27th Conf. on Learning Theory (COLT)*, 2014.

[14] O. Besbes and A. Zeevi. Dynamic Pricing Without Knowing the Demand Function: Risk Bounds and Near-Optimal Algorithms. *Operations Research*, 57(6):1407–1420, 2009.

[15] O. Besbes and A. Zeevi. On the Minimax Complexity of Pricing in a Changing Environment. *Operations Research*, 59:66–79, 2011.

[16] O. Besbes and A. Zeevi. Blind Network Revenue Management. *Operations Research*, 60:1520–1536, 2012.

[17] O. Besbes and A. Zeevi. On the Surprising Sufficiency of Linear Models for Dynamic Pricing with Demand Learning. 2013. Working paper, Columbia University.

[18] Omar Besbes and Costis Maglaras. Dynamic pricing with financial milestones: Feedback-form policies. *Management Science*, 58(9):1715–1731, 2012.

[19] Avrim Blum, Vijay Kumar, Atri Rudra, and Felix Wu. Online learning in online auctions. In *14th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 202–204, 2003.

[20] J. Broder and P. Rusmevichientong. Dynamic Pricing under a General Parametric Choice Model. *Operations Research*, 60:965–980, 2012.

[21] Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvari. Online Optimization in X-Armed Bandits. *J. of Machine Learning Research (JMLR)*, 12:1587–1627, 2011.

[22] W.L. Cooper, T. Homem de Mello, and A. J. Kleywegt. Models of the spiral-down effect in revenue management. *Operations Research*, 54:968–987, 2006.

[23] A. den Boer. Dynamic Pricing with Multiple Products and Partially Specified Demand Distribution. 2013. Working Paper, CWI, Amsterdam, The Netherlands.

[24] A. den Boer and B. Zwart. Simultaneously Learning and Optimizing using Controlled Variance Pricing. 2011. Working Paper, CWI, Amsterdam, The Netherlands.

[25] Miroslav Dudík, Daniel Hsu, Satyen Kale, Nikos Karampatziakis, John Langford, Lev Reyzin, and Tong Zhang. Efficient optimal leanring for contextual bandits. In *27th Conf. on Uncertainty in Artificial Intelligence (UAI)*, 2011.

[26] D. Easley and N. M. Kiefer. Controlling a Stochastic Process with Unknown Parameters. *Econometrica*, 56(5):1045–1064, 1988.

[27] S. S. Eren and C. Maglaras. Monopoly Pricing with Limited Demand Information. *Journal of Revenue and Pricing Management*, 9:23–48, 2010.

[28] V. F. Farias and B. van Roy. Dynamic Pricing with a Prior on Market Response. *Operations Research*, 58(1):16–29, 2010.

[29] A. Goldenshluger and A. Zeevi. Woodroofe's One-armed Bandit Problem Revisited. *Annals of Applied Probability*, 19(4):1603–1633, 2009.

[30] A. Goldenshluger and A. Zeevi. On a Linear Response Bandit Problem. *Stochastic Systems*, 3:230–261, 2013.

[31] Elad Hazan and Satyen Kale. Better algorithms for benign bandits. In *20th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 38–47, 2009.

[32] Elad Hazan and Nimrod Megiddo. Online Learning with Prior Information. In *20th Conf. on Learning Theory (COLT)*, pages 499–513, 2007.

[33] Chien-Ju Ho, Aleksandrs Slivkins, and Jennifer Wortman Vaughan. Adaptive contract design for crowdsourcing markets: Bandit algorithms for repeated principal-agent problems. In *15th ACM Conf. on Economics and Computation (EC)*, 2014. To appear in *J. of Artificial Intelligence Research*.

[34] G. Keller and S. Rady. Optimal Experimentation in a Changing Environment. *The Review of Economic Studies*, 66(3):475–507, 1999.

[35] G. Keller and S. Rady. Strategic Experimentation with Poisson Bandits. *Theoretical Economics*, 5:545–567, 2010.

[36] N. Bora Keskin and Assaf J. Zeevi. Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research*, 62(5), 2014.

[37] N.B. Keskin and A. Zeevi. Chasing Demand: Learning and Earning in a Changing Environment. 2013. Working paper, submitted to *Mathematics of Operation Research*.

[38] Robert Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *18th Advances in Neural Information Processing Systems (NIPS)*, 2004.

[39] Robert Kleinberg and Tom Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *44th IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 594–605, 2003.

[40] Robert Kleinberg and Aleksandrs Slivkins. Sharp dichotomies for regret minimization in metric spaces. In *21st ACM-SIAM Symp. on Discrete Algorithms (SODA)*, 2010.

[41] Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Multi-armed bandits in metric spaces. In *40th ACM Symp. on Theory of Computing (STOC)*, pages 681–690, 2008.

[42] Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Bandits and experts in metric spaces. Working paper, published at `http://arxiv.org/abs/1312.1277`, 2015. Merged and revised version of conference papers in *ACM STOC 2008* and *ACM-SIAM SODA 2010*.

[43] T. L. Lai and H. Robbins. Iterated Least Squares in Multiperiod Control. *Advances in Applied Mathematics*, 3(1):50–73, 1982.

[44] John Langford and Tong Zhang. The Epoch-Greedy Algorithm for Contextual Multi-armed Bandits. In *21st Advances in Neural Information Processing Systems (NIPS)*, 2007.

[45] Constantinos Maglaras. Dynamic pricing strategies for multi-product revenue management problems, 2010. Encyclopedia of Operations Research.

[46] Constantinos Maglaras and Joern Meissner. Dynamic Pricing Strategies for Multiproduct Revenue Management Problems. *Manufacturing & Service Operations Management*, 8(2):136–148, 2006.

[47] Odalric-Ambrym Maillard and Rémi Munos. Online Learning in Adversarial Lipschitz Environments. In *European Conf. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, pages 305–320, 2010.

[48] M. Rothschild. A Two-armed Bandit Theory of Market Pricing. *Journal of Economic Theory*, 9(2):185–202, 1974.

[49] Denis Sauré and Assaf Zeevi. Optimal dynamic assortment planning with demand learning. *Manufacturing & Service Operations Management*, 15(3):387–404, 2013.

[50] Adish Singla and Andreas Krause. Truthful incentives in crowdsourcing tasks using regret minimization mechanisms. In *22nd Intl. World Wide Web Conf. (WWW)*, pages 1167–1178, 2013.

[51] Aleksandrs Slivkins. Contextual bandits with similarity information. *J. of Machine Learning Research (JMLR)*, 15(1):2533–2568, 2014. Preliminary version in *COLT 2011*.

[52] Aleksandrs Slivkins and Jennifer Wortman Vaughan. Online decision making in crowdsourcing markets: Theoretical challenges. *SIGecom Exchanges*, 12(2), December 2013.

[53] S. Deng Wang, Z. and Y. Ye. Close the Gaps: A Learning-while-Doing Algorithm for a Class of Single-Product Revenue Management Problems. 2013. Working paper, Stanford University.

[54] Zizhuo Wang, Shiming Deng, and Yinyu Ye. Close the gaps: A learning-while-doing algorithm for single-product revenue management problems. *Operations Research*, 62(2):318–331, 2014.