Faculty of Economics,
University of Porto

Efficiency Measurement in Discrete Choice Models with and without Strategic Interaction

PhD. Thesis in Economics

Jorge Alves Pereira Cerdeira e Sousa

Supervised by: Elvira Maria de Sousa Silva

February 2012

Acknowledgments

First, I thank my supervisor, Elvira Silva, for all the support, encouragement, sound advice and enthusiasm throughout these years. Without her friendship, good teaching and interesting ideas, this thesis would not have materialized.

I would also like to express my gratitude to Victor Aguirregabiria, Ariel Pakes and Kenneth Train for all their help, which was decisive for my understanding of Discrete Choice Methods with Simulation and Empirical Industrial Organization. Thanks also go to Andrew Ching, Susumu Imai, Masakazu Ishihara and Pedro Mira for helpful hints on Dynamic Discrete Choice Models and to Natércia Fortuna for helpful comments on earlier versions of some of the essays.

I thank my friends for all the encouragement and support during this period. I am particularly grateful to Patrícia Carvalhido, Vasco Leite, Mónica Meireles, Luís Pina Rebelo and Nuno Torres for our fruitful conversations.

Some essays of this thesis have been presented at several seminars and conferences, namely at the XII European Workshop on Efficiency and Productivity Analysis, XXV Jornadas de Economía Industrial, 5th Annual International Symposium on Economic Theory, Policy and Applications, North American Productivity Workshop VI, CEF.UP and Faculty of Economics, University of Porto. I thank to all the participants and discussants for helpful comments.

This research has been funded by a PhD scholarship from the Portuguese Science Foundation (*Fundação para a Ciência e Tecnologia* - FCT), to which I express my gratitude.

Last, but definitely not least, I thank my family for always being there for me.

Summary

Efficiency measurement has been analyzed considering that firms make decisions over continuous variables. While this may be an adequate representation of firms' behaviour in many cases, there are some situations in which firms do not make decisions over continuous variables, but rather over discrete variables. In this thesis, we develop a simple approach to evaluate firm efficiency when discrete choice is considered.

In the first essay, entitled "Efficiency Measurement in a Dynamic Discrete Choice Model: An Application to Inventory Behaviour", we analyze a dynamic discrete choice inventory model and develop a measure of dynamic profit efficiency at the product level. We allow for the existence of product heterogeneity as well as efficiency heterogeneity across products and apply the model to data on a Portuguese firm. The results suggest that, on average, the firm obtains 81,6% of the maximum profit associated with a given product.

The second essay is entitled "Dynamic Efficiency and Machine Replacement: A Discrete Choice Approach" and presents a dynamic machine replacement model to analyze efficiency at the firm level. The model is applied to Portuguese data on 290 manufacturing firms from 2001 to 2008. The results indicate that the estimated inefficiency is very similar for big and small firms as actual costs are, on average, 1.41 and 1.42 times their minimum costs, respectively.

In the third essay, entitled "Efficiency Measurement in Dynamic Discrete Choice Processes", we generalize our approach, developing an efficiency measure for a dynamic discrete choice model in which the payoff function may be the cost function, the revenue function or the profit function. We also extend our efficiency measure to Empirical Games, that is, discrete choice models in which there is strategic interaction among firms. We illustrate the approach with the estimation of a dynamic empirical game for the Portuguese banking industry between 2002 and 2009.

Contents

List of Tables

List of Figures

Preamble

The productive efficiency measurement literature began in the fifties with the work by Koopmans (1951), Debreu (1951) and Farrell (1957). Koopmans (1951) proposes a definition of technical efficiency: an input-output vector is technically efficient if and only if increasing any output or decreasing any input is possible only by decreasing some other output or increasing some other input. Debreu (1951) proposes the first input-oriented technical efficiency measure, called the coefficient of resource utilization, which is the first radial measure of technical efficiency. Farrell (1957) presents the first empirical study on productive efficiency, applied to the USA agricultural sector, and decomposes cost efficiency into technical efficiency and allocative efficiency. The study by Farrell (1957) was important for the development of other measures of economic efficiency and the corresponding decompositions (e.g., see Färe *et al.* (1983, 1985, 1994)).

Nevertheless, the radial distance functions developed by Shephard (1953, 1970) play also an important role in the development of the literature on efficiency measurement. Shephard (1953) develops the input radial distance function and shows that this function can represent fully a production technology with multiple outputs and multiple inputs. The inverse of the radial input distance function is equal to the input-oriented radial technical efficiency, also called the Debreu-Farrell technical efficiency measure. Similarly, the output radial distance function developed by Shephard (1970) is equal to the inverse of the output-oriented technical efficiency measure (e.g., Färe *et al.* (1983, 1985, 1994)). The relationship between the radial distance functions and the technical efficiency measures is crucial for the development of the literature on the efficiency measurement (Kumbhakar and Lovell (2000)).

The theoretical literature on the productive efficiency measurement develops mainly on the basis of the radial distance functions. Several generalizations of the radial measures of efficiency have emerged in the literature (e.g., Färe and Lovell (1978), Deprins and Simar (1983), Färe *et al.* (1985), Briec (1997), Bogetoft and Hougaard (1998), Chambers *et al.* (1996, 1998), Färe and Grosskopf (2000a, 2000b), Chavas and Cox (1999), Halme *et al.* (1999)). In particular, the directional inefficiency measures, constructed on the basis of the directional distance functions, have an important role in recent developments on the efficiency (and productivity) measurement literature. Chambers *et al.* (1996, 1998) explore the work by Luenberger (1992, 1995) and propose the directional distance functions that allow the construction of directional measures of technical and economic inefficiency. The directional inefficiency measures are an important contribution to the measurement of productive efficiency for three main reasons. First, the input (output) directional measures do not impose a proportional change in the inputs (outputs) (Chambers *et al.* (1996)). In contrast, the radial input (output) measures, constructed on the basis of the radial input (output) distance functions, impose a proportional change in the inputs (outputs). Second, the technology directional distance function allows simultaneous adjustments in the inputs and outputs. In contrast, the radial functions only allow adjustments in the inputs or the outputs, but not in both. The directional technology distance function, which is dual to the profit function, allows the construction of a profit inefficiency measure and its decomposition in a parsimonious way. Profit has an additive structure, a feature that is shared with the directional distance function but not the radial distance functions, which are multiplicative in form (Chambers *et al.* (1998)). Third, the directional distance functions provide a simple framework to model and measure performance when there is joint production of good (desirable) and bad (undesirable) outputs. The directional inefficiency measures has been used recently in several empirical studies in different sectors (e.g., Färe *et al.* (2004), Färe *et al.* (2005), Glass *et al.* (2006)). For an overview of the directional inefficiency measures, see Färe and Grosskopf (2004).

The literature on efficiency measurement has been developed mainly in the context of the static theory of the firm. Recently, some studies develop productive efficiency measures in the context of intertemporal models of the firm (e.g., Sengupta (1995), Färe and Grosskopf (1996), Lasserre and Ouellette (1999), Nemoto and Goto (1999, 2003), Silva and Stefanou (2007), Rungsuriyawiboon and Stefanou (2007), Ouellette and Yan (2008), Silva and Lansink (2009)). Most of these research efforts are developed in the context of the adjustment-cost model of the firm. The dynamic models in Nemoto and Goto (1999, 2003) are constructed on the basis of a production possibility set defined in terms of variable inputs, quasi-fixed factors and outputs, where stocks of the quasi-fixed factors at the end of each period are treated as outputs while the stocks of these factors at the beginning of each period as treated as inputs. The dynamic factors (i.e., the change in the level of the quasi-fixed factors) are not explicitly modeled in the firm's production technology. In the dynamic models constructed in Silva and Stefanou (2007) and Ouellette and Yan (2008), the dynamic factors are explicitly incorporated in the firm's production technology. Silva and Stefanou (2007) considers that investment decisions are irreversible and develop hyperbolic input-oriented dynamic efficiency measures in the long- and short-run; Ouellette and Yan (2008) consider the possibility of investment and disinvestment and focus on the efficiency of variable inputs. Silva and Lansink (2009) establishes duality between a directional input distance function, defined within the adjustment-cost model of the firm, and the value function of the intertemporal cost minimization problem and construct directional input efficiency measures based on this dual relation. Serra *et al.* (2011) applies those measures to a sample of Dutch dairy farms observed from 1995-2005.

Efficiency measurement in either static or dynamic environments has been developed considering that firms make decisions over continuous variables. There are however some situations in which firms do not make decisions over continuous variables, but rather over discrete variables. Examples include decisions on product ordering, patent renewal, machine replacement, price changes, among others. The theoretical framework based on continuous choices may not be an adequate representation of the firm's behavior when discrete choice decisions are considered. In fact, these decision processes are more properly analyzed using the discrete choice framework.

Since the seminal work of McFadden (1978) on static discrete choice models, there has been a wide variety of applications using both analytical and simulation techniques in a static environment (e.g., McFadden and Train (1996), Train (1999), Hensher (2001), Small and Rosen (1981)). The extension of the framework to a dynamic environment is considered in Pakes (1986) and Rust (1987) with the analysis of Markov decision problems. Pakes (1986) uses simulation techniques to estimate a dynamic programming discrete choice model and uses the model to estimate the value of European patents. Rust (1987) proposes a nested fixed point (NFXP) algorithm, which combines a "outer algorithm" and a "inner algorithm". The first one searches for the roots of the likelihood function while the second solves the dynamic programming problem for each possible value of the structural parameters. Rust (1987) applies the estimation method to a machine replacement problem.

Since the studies by Pakes (1986) and Rust (1987), there have been several applications that use the NFXP algorithm in different areas of microeconomics. For instance, Sturm (1991), Das (1992), Kennet (1994) and Rust and Rothwell (1995) analyze investment models of machine replacement; Rust and Phelan (1997) and Kalstrom *et al.* (2004) investigate retirement from labor force; Gilleskie (1998) studies visits to a doctor during periods of illness; Ahn (1995) considers a model of fertility; and Kennan and Walker (2005) analyzes a model of migration decisions.

Though very intuitive and efficient, the NFXP algorithm is computationally very demanding, since the dynamic programming problem has to be solved in every iteration (Rust (1987)). This drawback has forced researchers to choose parsimonious specifications for the payoff functions and has limited the inclusion of sources of heterogeneity across agents in the models. The work of Hotz and Miller (1993) has been crucial to alleviate the computational burden associated with the estimation of dynamic discrete choice models. Hotz and Miller (1993) proves that there is an invertible mapping relating conditional choice probabilities and value functions. As a result, nonparametric estimates of choice probabilities can be used to recover the value functions without having to solve the dynamic programming problem. The method has been used in a variety of applications (e.g., Slade (1998), Aguirregabiria (1999), Rota (2004)).

While the Hotz and Miller (1993)'s estimator reduces the computational burden associated with the estimation of the model, it implies a loss on (both finite sample and asymptotic) efficiency (see the discussion on identification and properties of estimators in Rust (1994), Magnac and Thesmar (2002) and Aguirregabiria (2010)). In order to overcome this problem, Aguirregabiria and Mira (2002) suggests the Nested Pseudo Likelihood (NPL) method, which is a recursive method that iterates on conditional choice probabilities and the parameters that maximize the likelihood function. Like the Hotz and Miller (1993)'s estimator, the NPL method implies less computational burden than the NFXP algorithm, but the NPL estimator is asymptotically equivalent to the Maximum Likelihood Estimator, which is an efficient estimator (Aguirregabiria and Mira (2002)). Applications of the NPL method include Aguirregabiria and Alonso-Borrego (1999) on labor demand, Sanchez-Mangas (2002) and Lorincz (2005) on investment and machine replacement models, De Pinto and Nelson (2008) on land use and deforestation, among others.

The use of these computationally less intensive methods has allowed researchers to use more flexible specifications in the models. Yet, another concern regarding dynamic discrete choice models is related to the inclusion of sources of heterogeneity across agents in the model. To deal with this issue, several applications include the finite mixture framework (Heckman and Singer (1984)) in the dynamic discrete choice model. In these applications, agents are assumed to belong to a population with a fixed, known number of "types of agents" and the estimation procedure is adjusted to include the probability associated with the agent being of a given "type" (e.g., see Carro and Mira (2006), Mira (2007), Blau and Gilleskie (2008) and Arcidiacono *et al.* (2007)). The study by Imai *et al.* (2009) contributes significantly to overcome the concerns regarding heterogeneity in dynamic discrete choice models. Rather than using finite mixture models, Imai *et al.* (2009) employs a more flexible approach and shows how to easily estimate dynamic models with random parameters. Imai *et al.* (2009) proposes a Bayesian estimation method that allows simultaneously for the solution of the dynamic programming problem and the estimation of the parameters. This method includes two steps in each iteration: one step solves the dynamic programming model and the other one employs the Markov Chain Monte Carlo algorithm to draw values from the posterior distributions of the parameters. The inclusion of random parameters in dynamic discrete choice models may imply heavy computational burden when we consider classical maximum likelihood procedures (e.g., see Train (2009)). However, Imai *et al.* (2009)'s experiments show that with their Bayesian method it is possible to estimate a flexible dynamic discrete choice model that accounts for heterogeneity across agents without too much computational burden. Applications of Imai *et al.* (2009) include Norets (2008, 2009), Osborne (2011) and Ching *et al.* (2010), among others.

Recently, there have been some efforts to extend the dynamic discrete choice framework to dynamic Empirical Games, that is, discrete choice models in which there is strategic interaction among agents, meaning that each agent's payoff is affected by the other agents' decisions (e.g., Pesendorfer and Schmidt-Dengler (2008), Pakes *et al.* (2007), Bajari *et al.* (2007), Aguirregabiria and Mira (2007)). While strategic interaction has always been considered an important issue, the estimation of the dynamic model requires the agent's choice to be the best response to the other agents' choices. Yet, an estimation method that performs an optimization routine with such a constraint is computationally very demanding (e.g., see Aguirregabiria and Mira (2010)). Extensions of the Hotz and Miller (1993)'s estimator to dynamic games have proven to be successful. Jofre-Bonet and Pesendorfer (2003) uses an extension of the Hotz and Miller (1993)'s estimator to analyze a dynamic auction model and Pesendorfer and Schmidt-Dengler (2008) analyzes identification and estimation of dynamic discrete games. Pakes *et al.* (2007), Bajari *et al.* (2007) and Aguirregabiria and Mira (2007) propose different methods to estimate the model using extensions of the Hotz and Miller (1993)'s estimator, avoiding in this way the computation of the dynamic programming problem. All these methods consider that agents' choice probabilities reflect agents' beliefs about the behaviour of other agents, meaning that it is possible to use agents' conditional choice probabilities as agents' best responses, which significantly reduces the complexity of the problem (e.g., see Aguirregabiria and Mira (2010)). Pakes *et al.* (2007) proposes a Generalized Method of Moments (GMM) estimator; Bajari *et al.* (2007) develops a moment inequality estimator that can be applied to models where the parameters are either point-identified or set-identified; Aguirregabiria and Mira (2007) proposes a likelihood-based method - a Nested Pseudo Likelihood method - to estimate the parameters in the model. These methods have been used in several applications: Pakes *et al.* (2007)'s GMM estimator is applied in Dunne *et al.* (2006) to investigate firms' entry and exit choices in the USA; applications of Bajari *et al.* (2007)'s estimator include Ryan (2011), Holmes (2011), Beresteanu and Ellickson (2005) and Sweeting (2007); Aguirregabiria and Mira (2007)'s Nested Pseudo Likelihood estimator is used in Collard-Wexler (2007), Kano (2006), Aguirregabiria *et al.* (2007) and Aguirregabiria and Ho (2011), among others.

In these estimation methods for Empirical Games it is considered that agents' choice probabilities reflect agents' beliefs about the behaviour of other agents (see Pakes *et al.* (2007), Bajari *et al.* (2007) and Aguirregabiria and Mira (2007)). However, there are situations in which we would expect that agents face some strategic uncertainty (e.g., see Besanko *et al.* (2010)). This is likely the case, for instance, when we consider oligopoly markets - in which firms are very secretive about their strategies - or markets affected by policy changes - in which it is reasonable to believe that firms do not know the impact of a policy change in other firms' behaviour. To deal with this issue, Aguirregabiria and Magesan (2010) extends the NPL method of Aguirregabiria and Mira (2007) and considers a model with 2 firms and 2 choices in which firms' beliefs about the behaviour of the other firm may not be in equilibrium, that is, they do not represent actual behaviour of the other firm.

We use the dynamic discrete choice framework to analyze efficiency at the micro level. Like some of the papers on dynamic efficiency mentioned above, we consider that firms attempt to maximize intertemporal payoffs (e.g., Lasserre and Ouellette (1999), Nemoto and Goto (1999, 2003), Silva and Stefanou (2007), Ouellette and Yan (2008), Silva and Lansink (2009)). Also, we consider that inefficiency arises whenever firms deviate from their optimal, forward-looking, rational choices, therefore getting an actual payoff smaller than the optimal payoff. In contrast with the other studies, we explicitly assume that firms make decisions over discrete variables.

The thesis comprises three essays. In the first essay, entitled "Efficiency Measurement in a Dynamic Discrete Choice Model: An Application to Inventory Behaviour", we analyze a dynamic discrete choice inventory model - in which a firm decides whether to order or not some products in each time period - and develop a measure of dynamic profit efficiency at the product level. The model allows for the existence of product heterogeneity as well as efficiency heterogeneity across products by including random coefficients in the analysis. We use a dataset with weekly information on prices, sales, orders and stocks for a Portuguese firm from January 2008 to June 2009 and estimate the model with a two-stage approach. In the first stage, we provide nonparametric estimates of the transition probabilities of the state variables. In the second stage, we use the Bayesian estimation method proposed by Imai *et al.* (2009) to estimate the parameters of the model. The results suggest that the average product efficiency is around 81,6%, implying that, on average, the firm obtains 81,6% of the maximum profit associated with a given product. Also, the counterfactuals results indicate that the actual decisions diverge from the optimal decisions in at least 14.88% of the decisions and that volatility in sales explains a significant part of the estimated inefficiency.

The second essay is entitled “Dynamic Efficiency and Machine Replacement: A Discrete Choice Approach”. In this essay, we use a dynamic discrete choice framework to analyze efficiency at the firm level. Specifically, we analyze a machine replacement model - in which firms have to decide, in each period, whether to replace their machines or not so that their intertemporal costs are minimized - and introduce a measure of dynamic cost efficiency in this model. In the model, the structural parameters are random parameters to allow for the existence of firm heterogeneity as well as efficiency heterogeneity across firms. We use a dataset with yearly information for 290 Portuguese manufacturing firms from 2001 to 2008 and estimate the model for all the firms in our dataset, as well as for subsamples on big and small firms. The model is estimated using the Bayesian estimation method proposed by Imai *et al.* (2009). The results indicate that the estimated inefficiency is very similar for big and small firms: big and small firms’ actual costs are, on average, 1.41 and 1.42 times their minimum costs, respectively. Our estimation results also reveal that different types of firms have different cost structures. While big firms have, on average, significant maintenance costs and negligible learning costs, small firms bear important learning costs and statistically insignificant maintenance costs. Also, the counterfactuals results suggest that big firms should have done more replacements and small firms should have replaced less in order to be fully efficient.

The third essay, entitled “Efficiency Measurement in Dynamic Discrete Choice Processes”, generalizes our approach, developing an efficiency measure for a general dynamic discrete choice model - known as a Single-Agent Dynamic Model - in which the payoff function may be the cost function, the revenue function or the profit function. We also extend our efficiency measure to dynamic Empirical Games. We describe the Nested Pseudo Likelihood (NPL) Method proposed by Aguirregabiria and Mira (2002, 2007), which can be used to estimate Single-Agent Models and a particular class of Empirical Games. In addition, we generalize the identification results of Aguirregabiria and Magesan (2010), which uses an extension of the NPL method to estimate more general Empirical Games. The efficiency measure is illustrated with the estimation of a dynamic empirical game for the Portuguese banking industry between 2002 and 2009.

The results indicate that banks are, on average, revenue efficient, though the interaction among banks has different impacts for each service.

We innovate on some aspects with this thesis. First, we contribute to the literature on efficiency measurement by developing an unified approach to evaluate firm efficiency when discrete choice is considered. In fact, our approach can be applied to both static or dynamic models, with or without strategic interaction among firms, and it remains valid regardless of whether we are considering a cost, revenue or profit approach. Second, this thesis also offers an interesting methodological contribution to the literature on the econometrics of games. In essay 3, we extend the results on identification and estimation of dynamic games by considering a framework with an arbitrary (finite) number of firms, multinomial choice and firms' beliefs about the behaviour of other firms that may not be in equilibrium.

References

[1] Aguirregabiria, V. (1999), "The Dynamics of Markups and Inventories in Retailing Firms", Review of Economic Studies, 66, 275-308.

[2] Aguirregabiria, V. (2010), "Another Look at the Identification of Dynamic Discrete Decision Processes: An Application to Retirement Behavior", Journal of Business and Economic Statistics, 28, 201-218.

[3] Aguirregabiria, V. and C. Alonso-Borrego (1999), "Labor Contracts and Flexibility: Evidence from a Labor Market Reform in Spain," manuscript, University of Toronto.

[4] Aguirregabiria, V. and C. Ho (2011), "A Dynamic Oligopoly Game of the US Airline Industry: Estimation and Policy Experiments", Journal of Econometrics, *forthcoming*.

[5] Aguirregabiria, V. and A. Magesan (2010), "Estimation of Dynamic Discrete Games When Players' Beliefs are Not in Equilibrium", Working Paper, University of Toronto.

[6] Aguirregabiria, V. and P. Mira (2002), "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models", Econometrica, 70, 1519-1544.

[7] Aguirregabiria, V. and P. Mira (2007), "Sequential Estimation of Dynamic Discrete Games", Econometrica, 75, 1-53

[8] Aguirregabiria, V. and P. Mira (2010), "Dynamic Discrete Choice Structural Models: A Survey", Journal of Econometrics, 156, 38-67.

[9] Aguirregabiria, V., P. Mira, and H. Roman (2007), "An Estimable Dynamic Model of Entry, Exit and Growth in Oligopoly Retail Markets", American Economic Review, 97, 449-454.

[10] Ahn, N. (1995), "Measuring the Value of Children by Sex and Age Using a Dynamic Programming Model", Review of Economic Studies, 62, 361-379.

[11] Arcidiacono, P., A. Khwaja, and L. Ouyang (2007), "Habit Persistence and Teen Sex: Could Increased Access to Contraception Have Unintended Consequences for Teen Pregnancies?", Manuscript, Duke University.

[12] Bajari, P., L. Benkard, and J. Levin (2007), "Estimating Dynamic Models of Imperfect Competition", Econometrica, 75, 1331-1370.

[13] Beresteanu, A. and P. Ellickson (2005), "The Dynamics of Retail Oligopoly", Manuscript, Department of Economics, Duke University.

[14] Besanko, D., U. Doraszelski, L. Lu, and M. Satterthwaite (2010), "On the Role of Demand and Strategic Uncertainty in Capacity Investment and Disinvestment Dynamics", International Journal of Industrial Organization, 28, 383-389.

[15] Blau, D. and D. Gilleskie (2008), "The Role of Retiree Health Insurance in the Employment Behavior of Older Males", International Economic Review, 49, 475-514.

[16] Bogetoft, P.and J. Hougaard (1998), "Efficiency Evaluations Based on Potential (Non-Proportional) Improvements", Journal of Productivity Analysis, 12, 233-247.

[17] Briec, W. (1997), "A Graph-Type Extension of Farrell Technical Efficiency Measure", Journal of Productivity Analysis, 8, 95-110.

[18] Carro, J. and P. Mira (2006), "A Dynamic Model of Contraceptive Choice of Spanish Couples", Journal of Applied Econometrics, 21, 955-980.

[19] Chambers, R., Y. Chung and R. Färe (1996), "Benefit and Distance Functions", Journal of Economic Theory, 70, 407-419.

[20] Chambers, R., Y. Chung and R. Färe (1998), "Profit, Directional Distance Functions, and Nerlovian Efficiency", Journal of Optimization Theory, 98, 351-364.

[21] Chavas, J. and T. Cox (1999), “A Generalized Distance Function and the Analysis of Production Efficiency”, Southern Economic Journal, 66, 294-318.

[22] Ching, A., S. Imai, M. Ishihara, and N. Jain (2010), “A Guide to Bayesian Estimation of Dynamic Discrete Choice Models with an Application to a Store-Level Reward Program”, Available at SSRN: http://ssrn.com/abstract=1398444.

[23] Collard-Wexler, A. (2006), “Productivity Dispersion and Plant Selection in the Ready-Mix Concrete Industry”, Manuscript, New York University.

[24] Das, S. (1992), “A Micro-Econometric Model of Capital Utilization and Retirement: The Case of the U.S. Cement Industry”, Review of Economic Studies, 59, 277-297.

[25] De Pinto, A. and G. Nelson (2008), “Land Use Change with Spatially Explicit Data: A Dynamic Approach”, Environmental and Resource Economics, 43, 209-229.

[26] Debreu, G. (1951), “The Coefficient of Resource Utilization”, Econometrica, 22, 14-22.

[27] Deprins, D. and L. Simar (1983), “On Farrell Measures of Technical Efficiency”, Recherches Économiques de Louvain, 49, 123-137.

[28] Dunne, T., S. Klimek, M. Roberts, and Y. Xu (2006), “Entry and Exit in Geographic Markets”, Manuscript, Department of Economics, Pennsylvania State University.

[29] Färe, R. and S. Grosskopf (1996), “Intertemporal Production Frontiers with Dynamic DEA”, Kluwer Academic Publishers, Boston.

[30] Färe, R. and S. Grosskopf (2000a), “Theory and Application of Directional Distance Functions”, Journal of Productivity Analysis, 13, 93-103.

[31] Färe, R. and S. Grosskopf (2000b), “Notes on some Inequalities in Economics”, Economic Theory, 15, 227-233.

[32] Färe, R. and S. Grosskopf (2004), “New Directions: Efficiency and Productivity”, Kluwer Academic Publishers.

[33] Färe, R., S. Grosskopf, and K. Lovell (1983), “The Structure of Technical Efficiency”, Scandinavian Journal of Economics, 85, 181-190.

[34] Färe, R., S. Grosskopf, and K. Lovell (1985), "The Measurement of Efficiency of Production", Kluwer Academic Publishers.

[35] Färe, R., S. Grosskopf, and K. Lovell (1994), "Production Frontiers", Cambridge University Press.

[36] Färe, R., S. Grosskopf, and W. Weber (2004), "The Effect of Risk-Based Capital Requirements on Profit Efficiency in Banking", Applied Economics, 36, 1731-1743.

[37] Färe, R., S. Grosskopf, D. Noh, and W. Weber (2005), "Characteristics of a Polluting Technology: Theory and Practice", Journal of Econometrics, 126, 469-492.

[38] Färe, R. and K. Lovell (1978), "Measuring the Technical Efficiency of Production", Journal of Economic Theory, 19, 150-162.

[39] Farrell, M. J. (1957), "The Measurement of Productive Efficiency", Journal of the Royal Statistical Society, Series A, General, 120, 253-281.

[40] Gilleskie, D. (1998), "A Dynamic Stochastic Model of Medical Care Use and Work Absence", Econometrica, 66, 1-45.

[41] Glass, J., G. McCallion, D. McKillop and K. Stringer (2006), "A 'Technically Level Playing-field' Profit Efficiency Analysis of Enforced Competition between Publicly Funded Institutions", European Economic Review, 50, 1601-1626.

[42] Halme, M., T. Joro, P. Korhonen, S. Salo, and J. Wallenius (1999), "A Value Efficiency Approach to Incorporating Preference Information in Data Envelopment Analysis", Management Science, 45, 103-115.

[43] Heckman, J. and B. Singer (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Economic Models for Duration Data", Econometrica, 52, 271-320.

[44] Hensher, D. (2001), "The Valuation of Commuter Travel Time Savings for Car Drivers in New Zeland: Evaluating Alternative Model Specifications, Transportation, 28, 101-118.

[45] Holmes, T. (2011), "The Diffusion Of Wal-Mart And Economies of Density", Econometrica, 79, 253-302.

[46] Hotz, J. and R. Miller (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models", Review of Economic Studies, 60, 497-529.

[47] Imai, S., N. Jain and A. Ching (2009), “Bayesian Estimation of Dynamic Discrete Choice Models”, Econometrica, 77, 1865-1899.

[48] Jofre-Bonet, M. and M. Pesendorfer (2003), “Estimation of a Dynamic Auction Game”, Econometrica, 71, 1443-1489.

[49] Kano, K. (2006), “Menu Costs, Strategic Interactions and Retail Price Movements”, Manuscript, Queen’s University.

[50] Karlstrom, A., M. Palme, and I. Svensson (2004), “A Dynamic Programming Approach to Model the Retirement Behaviour of Blue-collar Workers in Sweden”, Journal of Applied Econometrics, 19, 795-807.

[51] Kennan, J. and J. Walker (2005), “The Effect Income on Individual Migration Decisions”, Manuscript, University of Wisconsin-Madison.

[52] Kennet, M. (1994), “A Structural Model of Aircraft Engine Maintenance”, Journal of Applied Econometrics, 9, 351-368

[53] Koopmans, T. C. (1951), “Analysis of Production as an Efficient Combination of Activities”, in: T. C. Koopmans (Eds.), Activity Analysis of Production and Allocation, Physica-Verlag, Heidelberg.

[54] Kumbhakar, S. and K. Lovell (2000), “Stochastic Frontier Analysis”, Cambridge University Press, New York.

[55] Lasserre, P. and P. Ouellette (1999), “Dynamic Factor Demands and Technology Measurement under Arbitrary Expectations”, Journal of Productivity Analysis, 11, 219-241.

[56] Lorincz, S. (2005), “Persistence Effects in a Dynamic Discrete Choice Model: Application to Low-End Computer Servers”, Discussion Papers 2005/10, Institute of Economics Hungarian Academy of Sciences.

[57] Luenberger, D. (1992), “Benefit Functions and Duality”, Journal of Mathematical Economics, 21, 461-481.

[58] Luenberger, D. (1995), “Microeconomic Theory”, McGraw-Hill, Boston.

[59] Magnac, T. and D. Thesmar (2002), “Identifying Dynamic Discrete Decision Processes”, Econometrica, 70, 801-816.

[60] McFadden, D. (1978), “Modelling the Choice of Residential Location”, in: A. Karlquist *et al.*, (Eds.), Spatial Interaction Theory and Planning Models, North Holland, Amsterdam, 75–96.

[61] McFadden, D. and K. Train (1996), “Consumers’ Evaluation of New Products: Learning from Self and Others”, Journal of Political Economy, 104, 683-703.

[62] Mira, P. (2007), “Uncertain Infant Mortality, Learning and Life-Cycle Fertility”, International Economic Review, 48, 809-846.

[63] Nemoto, J. and M. Goto (1999), “Dynamic Data Envelopment Analysis: Modeling Intertemporal Behavior of a Firm in the Presence of Productive Inefficiencies”, Economic Letters, 64, 51-56.

[64] Nemoto, J. and M. Goto (2003), “Measurement of Dynamic Efficiency in Production: An Application of Data Envelopment Analysis to Japanese Electric Utilities”, Journal of Productivity Analysis, 19, 191-210.

[65] Norets, A. (2008), “Implementation of Bayesian Inference in Dynamic Discrete Choice Models”, Manuscript, Princeton University.

[66] Norets, A. (2009), “Inference in Dynamic Discrete Choice Models with Serially Correlated Unobserved State Variables”, Econometrica, 77, 1665–1682.

[67] Osborne, M. (2011), “Consumer Learning, Switching Costs and Heterogeneity: A Structural Examination”, Quantitative Marketing and Economics, 9, 25-70.

[68] Ouellette, P. and L. Yan (2008), “Investment and Dynamic DEA”, Journal of Productivity Analysis, 29, 235-247.

[69] Pakes, A. (1986), “Patents as Options: Some Estimates of the Value of Holding European Patent Stocks”, Econometrica, 54, 755-784.

[70] Pakes, A., M. Ostrovsky and S. Berry (2007), “Simple Estimators for the Parameters of Discrete Dynamic Games (with entry/exit examples)”, Rand Journal of Economics, 38, 373-399.

[71] Pesendorfer, M. and P. Schmidt-Dengler (2008), “Asymptotic Least Squares Estimators for Dynamic Games”, Review of Economic Studies, 75, 901-928.

[72] Rota, P. (2004), “Estimating Labor Demand with Fixed Costs”, International Economic Review, 45, 25-48.

[73] Rungsuriyawiboon, S. and S. Stefanou (2007), "Dynamic Efficiency Estimation: An Application to US Electric Utilities", Journal of Business and Economic Statistics, 25, 226-238.

[74] Rust, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher", Econometrica, 55, 999-1033.

[75] Rust, J. (1994), "Structural Estimation of Markov Decision Processes", in: R. E. Engle and D. McFadden, (Eds.), Handbook of Econometrics, Vol. 4, North-Holland, Amsterdam, 3081-3143.

[76] Rust, J. and C. Phelan (1997), "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets", Econometrica, 65, 781-832.

[77] Rust, J. and Rothwell, G. (1995), "Optimal Response to a Shift in Regulatory Regime: The Case of the US Nuclear Power Industry", Journal of Applied Econometrics, 10, S75-S118.

[78] Ryan, S. (2011), "The Costs of Environmental Regulation in a Concentrated Industry", Econometrica, *forthcoming*.

[79] Sanchez-Mangas, R. (2002), "Pseudo-Maximum Likelihood Estimation of a Dynamic Structural Investment Model", Working Paper 02-62, Statistics and Econometrics Series, Universidad Carlos III de Madrid.

[80] Sengupta, J. (1995), "Dynamics of Data Envelopment Analysis", Kluwer Academic Publishers, Dordrecht, The Netherlands.

[81] Serra, T., A. O. Lansink and S. Stefanou (2011), "Measurement of Dynamic Efficiency, a Directional Distance Function Parametric Approach", American Journal of Agricultural Economics, 93, 756-767.

[82] Shephard, R. (1953), "Cost and Production Functions", Princeton University Press, Princeton.

[83] Shephard, R. (1970), "Theory of Cost and Production Functions, Princeton University Press, Princeton.

[84] Silva, E., and A. Lansink (2009), "Dynamic Efficiency Measurement: a Directional Distance Function Approach", Unpublished, Wageningen University, in review

[85] Silva, E. and S. Stefanou (2007), "Dynamic Efficiency Measurement: Theory and Application", American Journal of Agricultural Economics, 89, 398-419.

[86] Slade, M. (1998), “Optimal Pricing with Costly Adjustment: Evidence from Retail Grocery Stores”, Review of Economic Studies, 65, 87-108.

[87] Small, K. and H. Rosen (1981), “Applied Welfare Economics of Discrete Choice Models”, Econometrica, 49, 105-130.

[88] Sturm, R. (1991), “A Structural Economic Model of Operating Cycle Management in European Nuclear Power Plants, Manuscript, RAND Corporation.

[89] Sweeting, A. (2007), “Dynamic Product Repositioning in Differentiated Product Markets: The Case of Format Switching in the Commercial Radio Industry”, NBER Working Paper 13522.

[90] Train, K. (1999), “Mixed Logit Models for Recreation Demand”, in: J. Herriges and C. Kling (Eds.), Valuing Recreation and the Environment, Edward-Elgar, Northampton, MA.

[91] Train, K. (2009), “Discrete Choice Methods with Simulation”, Cambridge University Press, 2nd Edition.

Essay 1: Efficiency Measurement in a Dynamic Discrete Choice Model: An Application to Inventory Behaviour

Abstract

We analyze a dynamic programming inventory model where a firm decides over discrete variables and develop a measure of dynamic efficiency at the product level. We estimate the structural parameters using the Bayesian estimation method proposed by Imai *et al.* (2009), which allows simultaneously for the solution of the dynamic programming problem and the estimation of the parameters. Our counterfactual experiments indicate that actual decisions diverge from optimal decisions in at least 14.88% of the decisions and that volatility in sales is responsible for a significant part of the estimated inefficiency.

Keywords: Dynamic Programming Inventory Model, Dynamic Efficiency, Bayesian Estimation Methods

JEL Classification: C15, C25, C61, D21

1 Introduction

Dynamic efficiency measurement at the micro level has been developed in the context of models in which firms decide over continuous variables [e.g., Silva and Stefanou (2007), Nemoto and Goto (1999, 2003), Lasserre and Ouellette (1999), Ouellette and Yan (2008)]. There are, however, many situations in which the firm makes decisions over discrete rather than continuous variables (e.g., a firm decides whether to order or not some products in each period). In this paper, dynamic (profit) efficiency is investigated within a dynamic discrete choice structural model, where agents decide over discrete rather than continuous variables. We analyze a dynamic programming inventory model - in which a firm decides whether to order or not some products in each period - and develop a measure of dynamic efficiency at the product level.

Since the seminal work of Arrow *et al.* (1951) and Scarf (1959), dynamic models of inventory behaviour have been used to explain behaviour at a micro level (e.g., see Deneckere *et al.* (1996) and Hall and Rust (2000)), as well as to analyze business cycles, since recessions are, in general, associated with periods of inventory liquidations (e.g., see Blinder (1981) and Ramey and West (1999)). One of the frameworks that have been used to analyze inventory decisions is based on dynamic discrete choice structural models [e.g., Aguirregabiria (1999); Aguirregabiria and Mira (2010) presents an interesting survey on several applications of dynamic discrete choice models]. Aguirregabiria (1999) analyzes a discrete choice model of prices and inventory behaviour, providing insight on the dynamics of inventories and markups at a micro level. Using a dynamic discrete choice structural model, Aguirregabiria (1999) is able to perform counterfactual experiments to evaluate the impact of changes in ordering costs on firm's decisions.

In fact, the possibility of performing counterfactual experiments, that is, to evaluate the impact of a change in structural parameters on agent's decisions, is one of the major advantages of dynamic discrete choice structural models. In this type of models, agents are forward looking and maximize expected intertemporal payoffs. Under the principle of revealed preferences, we can use data on agent's choices and outcomes to estimate the structural parameters. These parameters are structural in the sense that they have precise economic meanings, representing agent's payoffs, preferences and beliefs about future events.

We use the dynamic discrete choice framework to develop a measure of dynamic efficiency in an inventory model with both variable and fixed ordering costs. The model also allows for the existence of product heterogeneity as well as efficiency heterogeneity across products by including random coefficients in the analysis. Using a dataset with weekly information on prices, sales, orders and stocks for a Portuguese firm from January 2008 to June 2009, we estimate the model with a two-stage approach. In the first stage, we provide nonparametric estimates of the transition probabilities of the state variables. In the second stage, we use the Bayesian estimation method proposed by Imai *et al.* (2009), which allows simultaneously for the solution of the dynamic programming problem and the estimation of the parameters. This method includes in each iteration two steps: one solves the dynamic programming model and the other employs the Markov Chain Monte Carlo (MCMC) algorithm to draw values from the posterior distributions of the parameters. We also perform counterfactual experiments to investigate what would be the firm's optimal choices and compare them with the actual choices. The counterfactuals results indicate that the actual decisions diverge from the optimal decisions in at least 14.88% of the decisions and that volatility in sales explains a significant part of the estimated inefficiency.

The paper is organized as follows. In Section 2, we present the dynamic programming inventory model. Section 3 discusses the efficiency measurement in our model. We present the estimation method in Section 4 and the data and estimation results are presented, respectively, in Sections 5 and 6. Our counterfactual experiments are discussed in Section 7. Section 8 concludes the paper.

2 The Inventory Model

In this Section, we present a discrete choice dynamic programming model where a multiproduct firm decides, in every period, whether to order or not each product. The firm buys its products in the wholesale market and sells the products to its clients. Time is discrete and indexed by t, $t = 0,...,\infty$, while products are indexed by i, $i = 1,..., N$. The decision variables, a_{it}, belong to a discrete, finite set $A = \{0, 1\}$, where $a_{it} = 1$ if the firm orders product i in time period t and $a_{it} = 0$ otherwise.

The firm's revenue for a given product depends on the quantity of the product sold as well as the sales price. We consider that when the firm makes its ordering decisions in each time period, the sales for that period are not known, so the firm bases its decisions on expected sales. We define p_{it} as the sales price of product i in period t and we denote expected sales in physical units by $E(y_{it})$. We also include in the firm's expected revenue an indicator function, $E(I\{t > t_{tax\,change}\})$, that captures the effect of a VAT (value added tax) change in expected sales.

Sales are assumed to be the minimum of inventories and demand. The firm cannot sell more than the demand for a product, but it is also possible that, in the case of a positive shock in the demand, the firm cannot satisfy it with the available stock of the product. In the spirit of Aguirregabiria (1999), we assume that expected sales are equal to

$$E(y_{it}) = d_{it} E\left(min\left\{ \frac{s_{it} + q_{it}}{d_{it}}, exp\{\phi_t\} \right\} \right), \tag{1.1}$$

where q_{it} represents the quantity of product i ordered in period t, s_{it} is the stock at the beginning of period t, d_{it} is the expected demand for product i in period t and ϕ_t is an iid demand shock that is known by the firm only after it has made the order decision for that period. We consider the isoelastic expected demand $d_{it} = exp\{\eta\, log(p_{it})\}$, where η is the price elasticity of the expected demand.

In the event of a stock-out, the demand for the product is not entirely satisfied with the available stock of that product. In this case, the firm perceives a profit loss that results from the fact that some customers that cannot buy the product to the firm will buy it at alternative suppliers. We include in our model a parameter to account for the negative effect of a stock-out in profits and we specify this loss in profits as depending on the probability of a stock-out, $P\{y_{it} = s_{it} + q_{it}\}$. Given the definition of an indicator function ($I\{\xi\} = 1$ if ξ is true and $I\{\xi\} = 0$ otherwise), we can define $P\{y_{it} = s_{it} + q_{it}\} = E\left(I\{y_{it} = s_{it} + q_{it}\}\right)$.

The firm's costs for a given product include storage costs and costs with the acquisition of the product. We consider that the storage cost is linear in the quantity of the product stored, s_{it}. We define c_{it} as the wholesale price of product i at time period t. So, $c_{it}q_{it}$ represents the cost of the product that the firm bears when it orders q_{it} units of product i at time period t. In addition to this cost, the firm also faces some ordering costs due to the existence of transportation costs and other costs associated with the ordering process. We include in our model both a fixed and a variable components in the ordering costs.

Let us represent the observed state variables by x_{it}, $x_{it} = (p_{it}, c_{it}, s_{it})'$. We assume that the sales price and wholesale price follow exogenous first-order markov processes $f_p(p_{it+1} \mid p_{it})$ and $f_c(c_{it+1} \mid c_{it})$, respectively.[1] Also, the stock variable has the following transition law: $s_{it+1} = max\{0, s_{it} + q_{it} - y_{it}\}$.

The expected value of the current profit function for product i conditional on the decision variable a_{it} and the state variables x_{it} is given by

$$E_\pi(a_{it}, x_{it}) = p_{it}E\left(y_{it} \mid x_{it}, a_{it}\right)\left(1 + \lambda E\left(I\{t > t_{tax\,change}\} \mid x_{it}, a_{it}\right)\right)$$
$$-(c_{it} + \gamma_i)E(q_{it} \mid x_{it}, a_{it}) - \mu s_{it} - \omega_i E\left(I\{y_{it} = s_{it} + q_{it}\} \mid x_{it}, a_{it}\right) - \zeta_i a_{it}. \tag{1.2}$$

Parameters in the expected value of the current profit function include the measure of the effect of a VAT change in expected sales λ, the variable ordering cost γ_i, the unit storage cost μ, the negative effect in profits resulting from a stock-out ω_i and the fixed ordering cost ζ_i. While λ and μ are fixed parameters, γ_i, ω_i and ζ_i are random parameters to take into account product heterogeneity. Specifically, we assume a lognormal distribution for these random parameters: $ln\gamma_i \sim N(\overline{ln\,\gamma}, \sigma^2_{ln\,\gamma})$, $ln\omega_i \sim N(\overline{ln\,\omega}, \sigma^2_{ln\,\omega})$ and $ln\zeta_i \sim N(\overline{ln\,\zeta}, \sigma^2_{ln\,\zeta})$. Hereafter, we denote all parameters in the current profit function by θ.

Given the definition of $E_\pi(a_{it}, x_{it})$, we can write the current profit function for product i as

$$\pi(a_{it}, x_{it}, \varepsilon_{it}) = E_\pi(a_{it}, x_{it}) + \varepsilon_{it}(a_{it}), \tag{1.3}$$

[1]The firm operates in a market with a large number of firms so it is unlikely that changes in the firm's ordering decisions will lead to significant changes in prices. Therefore, it is reasonable to assume that c_{it} and p_{it} are exogenously determined.

where $\varepsilon_{it}(a_{it})$ represents the uncertainty that the econometrician has about the actual expected profits observable to the firm. We assume that ε_{it} is independent over time with type 1 extreme value distribution $G(\varepsilon_{it})$.

Given the state variables, the problem of the firm is to make decisions a_{it} in order to maximize the expected discounted flow of profits over time for product i

$$E\left(\sum_{t=0}^{\infty}\delta^{t}\,\pi(a_{it},x_{it},\varepsilon_{it})\right), \tag{1.4}$$

where $\delta\in(0,1)$ is the discount factor.

Let $\alpha(x_{it},\varepsilon_{it};\theta)$ and $V(x_{it},\varepsilon_{it};\theta)$ denote the optimal decision rule and the value function of the dynamic programming problem. We can obtain the value function in a recursive fashion as

$$V(x_{it},\varepsilon_{it};\theta)=\max_{a\in A}\left\{E_{\pi}(a,x_{it})+\varepsilon_{it}(a)+\delta E_{x,\varepsilon}\left[V(x_{it+1},\varepsilon_{it+1};\theta)\mid a,x_{it},\varepsilon_{it}\right]\right\}. \tag{1.5}$$

The optimal decision rule is $\alpha(x_{it},\varepsilon_{it};\theta)=arg\,max_{a\in A}\left\{v(a,x_{it},\varepsilon_{it};\theta)\right\}$, where for each $a\in A$,

$$v(a,x_{it},\varepsilon_{it};\theta)\equiv E_{\pi}(a,x_{it})+\varepsilon_{it}(a)+\delta E_{x,\varepsilon}\left[V(x_{it+1},\varepsilon_{it+1};\theta)\mid a,x_{it},\varepsilon_{it}\right]. \tag{1.6}$$

Given the structure of the current profit function and the assumptions made about the transition probabilities of the state variables, we can rewrite problem (1.5) using the concept of integrated Bellman equation (see Aguirregabiria and Mira (2010)),

$$\begin{aligned}\bar{V}(x_{it};\theta)&\equiv\int V(x_{it},\varepsilon_{it};\theta)\,dG(\varepsilon_{it})\\&=\int\max_{a\in A}\left\{E_{\pi}(a,x_{it})+\varepsilon_{it}(a)+\delta E_{x}\left[\bar{V}(x_{i,t+1};\theta)\mid a,x_{it}\right]\right\}dG(\varepsilon_{it}).\end{aligned} \tag{1.7}$$

In addition, define

$$\bar{v}(a,x_{it};\theta)\equiv E_{\pi}(a,x_{it})+\delta E_{x}\left[\bar{V}(x_{i,t+1};\theta)\mid a,x_{it}\right], \tag{1.8}$$

and denote the optimal rule by $\bar{\alpha}(x_{it},\varepsilon_{it};\theta)=arg\,max_{a\in A}\left\{\bar{v}(a,x_{it};\theta)+\varepsilon_{it}(a)\right\}$.

We define the Conditional Choice Probability (CCP), which is a component of the likelihood function, as

$$P(a \mid x;\theta) = \int I\left\{\bar{\alpha}(x,\varepsilon;\theta) = a\right\} dG(\varepsilon)$$

$$= \int I\left\{\bar{v}(a, x_{it};\theta) + \varepsilon_{it}(a) > \bar{v}(a', x_{it};\theta) + \varepsilon_{it}(a'),\ \forall a'\right\} dG(\varepsilon_{it}). \tag{1.9}$$

Given that ε_{it} is independent over time with type 1 extreme value distribution $G(\varepsilon_{it})$, equations (1.7) and (1.9) can be expressed as (see Aguirregabiria and Mira (2010))

$$\bar{V}(x_{it};\theta) = log\left(\sum_{a\in A} exp\left\{\bar{v}(a, x_{it};\theta)\right\}\right), \tag{1.10}$$

and

$$P(a \mid x;\theta) = \frac{exp\left\{\bar{v}(a, x_{it};\theta)\right\}}{\sum_{j=0}^{1} exp\left\{\bar{v}(a=j, x_{it};\theta)\right\}}. \tag{1.11}$$

Given (1.11), we have

$$P(a \mid x) = \int \frac{exp\left\{\bar{v}(a, x_{it};\theta)\right\}}{\sum_{j=0}^{1} exp\left\{\bar{v}(a=j, x_{it};\theta)\right\}}\, dF_{\theta}, \tag{1.12}$$

where F_{θ} represents the distribution of θ.

Having data on $i = 1, ..., N$ products during $t = 1, ..., T$ periods, we can define the (conditional) likelihood function as

$$L(\theta) = \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=0}^{1} P(a \mid x;\theta)^{I\{a=j\}}, \tag{1.13}$$

and so the (unconditional) likelihood function is defined by

$$L = \int \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=0}^{1} P(a \mid x;\theta)^{I\{a=j\}}\, dF_{\theta}. \tag{1.14}$$

3 The Efficiency Measure

Dynamic efficiency measurement at the micro level is still in an infancy stage. Dynamic efficiency has been evaluated within models where choices are represented by continuous variables (e.g., Silva and Stefanou (2007), Nemoto and Goto (1999, 2003), Lasserre and Ouellette (1999), Ouellette and Yan (2008), Serra *et al.* (2011)). For example, Silva and Stefanou (2007) develops a measure of dynamic cost efficiency within an adjustment-cost model of the firm where the firm's choices concern the amount of variable and dynamic factors in each time period. In contrast, we attempt to develop a product-specific profit efficiency measure in the context of the dynamic discrete choice structural model presented in the previous Section. The focus here is on the decision of ordering new deliveries with the firm having only two possible choices regarding each product in each period: choosing $a_{it} = 0$ or $a_{it} = 1$. Inefficiency arises whenever the firm actually deviates from its optimal, forward-looking, rational choices, therefore getting a profit smaller than the maximum profit.

A natural way to analyze profit efficiency in a model is to include an additive term in the current profit function that can only assume non-positive values (so that in the event of inefficiency actual profits are smaller than optimal profits). This is an approach that is in the spirit of the Stochastic Frontier Analysis (e.g., see Kumbhakar and Lovell (2000)). However, in a discrete choice model "only differences in utility matter" (see Train (2009)) and such an efficiency term will not be identifiable.[2]

The current profit function for product i in (1.3) represents the maximum profit that the firm obtains with product i. Allowing for inefficiency in the ordering decisions, the current profit function in (1.3) is redefined as

$$\pi(a_{it}, x_{it}, \varepsilon_{it}) = E_\pi(a_{it}, x_{it}) + \rho_i + \varepsilon_{it}(a_{it}),$$

where ρ_i represents the profit loss due to inefficiency. Because ρ_i is not identifiable, we define ρ_i as a function of several observable variables, that is, $\rho_i = (\Gamma(\beta_i) - 1)\, E_\pi(a_{it}, x_{it})$. Substituting ρ_i in the previous equation yields

$$\pi(a_{it}, x_{it}, \varepsilon_{it}) = \Gamma(\beta_i) \times E_\pi(a_{it}, x_{it}) + \varepsilon_{it}(a_{it}), \tag{1.15}$$

where $\Gamma(\beta_i)$ is defined as

$$\Gamma(\beta_i) = \begin{cases} \beta_i & if \;\; E_\pi(a_{it}, x_{it}) \geq 0 \\ 1/\beta_i & if \;\; E_\pi(a_{it}, x_{it}) < 0 \end{cases} \tag{1.16}$$

[2]Consider the CCP defined in (1.9) and denote the profit loss due to inefficiency by ρ_i. The term ρ_i will not be identifiable in our model because the CCP will be the same with or without the inclusion of ρ_i. In fact, the CCP in (9) is equal to the following CCP in which ρ_i is included:

$$P(a \mid x; \theta) = \int I\left\{\bar{v}(a, x_{it}; \theta) + \rho_i + \varepsilon_{it}(a) > \bar{v}(a', x_{it}; \theta) + \rho_i + \varepsilon_{it}(a'),\, \forall a'\right\} dG(\varepsilon_{it}).$$

and $0 < \beta_i \leq 1$.

The parameter β_i is the product-specific efficiency measure. If the firm's ordering choice regarding product i is optimal, then $\beta_i = 1$; if the firm's choice regarding product i is not optimal, $\beta_i < 1$ and the firm gets a profit which is smaller than the maximum profit. So, we are considering that, in the event of inefficiency, the firm only gets a fraction of the maximum (expected) profit for the product.

It is possible that some products in some time periods are associated with negative profits. In order to deal with the possible existence of negative profits, the function $\Gamma(\beta_i)$ assumes different values for positive and negative profits.[3] If $E_\pi(a_{it}, x_{it}) \geq 0$, then $0 < \Gamma(\beta_i) = \beta_i \leq 1$ and actual profit is equal to maximum profit only if $\beta_i = 1$; conversely, if $E_\pi(a_{it}, x_{it}) < 0$, then $\Gamma(\beta_i) = 1/\beta_i \geq 1$ and actual profit is less than or equal to maximum profit. In fact, if $E_\pi(a_{it}, x_{it}) \geq 0$, then, conditional on a_{it} and x_{it}, the (expected) maximum profit is $\frac{1}{\beta_i}$ times actual profits; if $E_\pi(a_{it}, x_{it}) < 0$, then, conditional on a_{it} and x_{it}, actual losses are $\frac{1}{\beta_i}$ times (expected) minimum losses. Note that our specification of the efficiency measure does not rule out the possibility of having $E_\pi(a_{it}, x_{it}) \geq 0$ and actual profit $\pi(a_{it}, x_{it}, \varepsilon_{it}) < 0$ as the difference in signals may be explained by a sufficiently low ε_{it}.

The profit function specification in (1.15) assumes no interaction between β_i and $\varepsilon_{it}(a_{it})$. This is a standard assumption in the literature on the Stochastic Frontier Analysis (e.g., see Kumbhakar and Lovell (2000)). In our model, the assumption of independence of β_i and $\varepsilon_{it}(a_{it})$ is crucial from an estimation perspective. In fact, if we did not consider such assumption, β_i would not be identifiable.[4]

We allow for efficiency heterogeneity across products by treating β_i as a random parameter. Specifically, we assume a truncated normal distribution for β_i: $\beta_i \sim TN_{[0,1]}(\overline{\beta^{\mathbf{T}}}, \sigma^2_{\beta^T})$, where TN stands for the truncated normal distribution and $[0,1]$ represents the truncation interval with the lower bound equal to zero and the upper bound equal to 1.

[3] Before the introduction of the directional technology distance function in the literature on static efficiency measurement (see Chambers *et al.* (1998)), which provides an additive measure of profit efficiency, there was no consensus on a profit efficiency measure (see Färe *et al.* (2004)). A measure of profit efficiency was, in general, defined as the ratio of actual profit over maximum profit. However, the ratio measure is not well-defined in the case of negative profits, leading to ad-hoc adjustments in this measure. For instance, Berger and Mester (1997) considers a ratio between actual and optimal profits where a positive constant is added to each firm's profit so that profits are always positive.

[4] Consider the CCP defined in (1.9). With an efficiency term affecting all the current profit function, the CCP would be defined as

$$P(a \mid x;\theta) = \int I\left\{\Gamma(\beta_i)\pi(a_{it}, x_{it}, \varepsilon_{it}) + \delta E_x\left[\bar{V}(.) \mid a, x_{it}\right] > \Gamma(\beta_i)\pi(a'_{it}, x_{it}, \varepsilon_{it}) + \delta E_x\left[\bar{V}(.) \mid a', x_{it}\right], \forall a'\right\} dG(\varepsilon_{it})$$

$$= \int I\left\{\pi(a_{it}, x_{it}, \varepsilon_{it}) + \delta^* E_x\left[\bar{V}(.) \mid a, x_{it}\right] > \pi(a'_{it}, x_{it}, \varepsilon_{it}) + \delta^* E_x\left[\bar{V}(.) \mid a', x_{it}\right], \forall a'\right\} dG(\varepsilon_{it})$$

and $\delta^* = \delta/\Gamma(\beta_i)$. But δ^* is not identifiable as the discount factor in this type of models is not identifiable (e.g., see Rust (1994), Magnac and Thesmar (2002)).

We estimate the model presented in Section 2 with the current profit function specified in (1.15) and the parameters of the density function of β_i (i.e., the average efficiency across products and the associated standard deviation).

4 Estimation Method

We use a two-stage approach to estimate the structural model. In the first stage, we estimate the transition probabilities of the state variables and the terms $E(q_{it} \mid x_{it}, a_{it})$, $E(y_{it} \mid x_{it}, a_{it})$, $E(I\{t > t_{tax\,change}\} \mid x_{it}, a_{it})$ and $E(I\{y_{it} = s_{it} + q_{it}\} \mid x_{it}, a_{it})$ using nonparametric methods (see appendix A and appendix B for details). In the second stage, we exploit the discrete choice decision to estimate the remaining parameters - conditional on the estimates obtained in the first stage - using the Bayesian estimation procedure suggested in Imai *et al.* (2009) and also analyzed in Ching *et al.* (2010). This estimation procedure involves the usage of Markov Chain Monte Carlo (MCMC) algorithms and Bayesian methods, which do not require the maximization of the likelihood function. Maximization of this function could be numerically difficult given the heterogeneity allowed in our model. Instead, the Bayesian procedure consists on specifying a prior for every parameter to be estimated and then drawing many values from the posterior distribution of the parameters conditional on the observed data.

Let us specify the priors and proposal distributions for the parameters. Recall that we have denoted all the parameters in the current profit function by θ. Let us define $\theta = (\theta_i, \theta_1)$, where $\theta_i = (\gamma_i, \omega_i, \zeta_i, \beta_i)$ and $\theta_1 = (\lambda, \mu)$. We do not include the discount factor, δ, because in this type of models δ is nonparametrically non-identified (see Rust (1994), Magnac and Thesmar (2002) for details). Therefore, we set $\delta = 0.985$ in the estimation process. In addition, we set $\bar{\theta} = (\overline{ln\,\gamma}, \overline{ln\,\omega}, \overline{ln\,\zeta}, \overline{\beta^T})$ and $\theta_\sigma = (\sigma_{ln\,\gamma}, \sigma_{ln\,\omega}, \sigma_{ln\,\zeta}, \sigma_{\beta^T})$. We specify a Normal prior for each term of $\bar{\theta}$ and an Inverted Gamma prior for each term of θ_σ. For the parameters in θ_1, we use a flat prior (i.e., we set the prior to be equal to 1). Also, we define a normal random-walk proposal distribution for each term of θ_1. The proposal distributions for θ_i have already been specified in Sections 2 and 3. Our goal is, therefore, to estimate the parameters in θ_1, $\bar{\theta}$ and θ_σ.

The posterior distribution of the parameters, $\Lambda(.)$, is defined as

$$\Lambda(\theta_i, \theta_1, \bar{\theta}, \theta_\sigma) \propto L(\theta_i, \theta_1) pd(\theta) k(\theta_1, \bar{\theta}, \theta_\sigma). \tag{1.17}$$

$\Lambda(.)$ is proportional to the expression in the right-hand side of (1.17), which depends on the likelihood function $L(\theta_i, \theta_1)$ defined in (1.13), the priors $k(\theta_1, \bar{\theta}, \theta_\sigma)$ and the proposal distributions $pd(\theta)$.

Our goal is, therefore, to repeatedly draw values from the posterior distribution (1.17), obtaining, in each iteration, one value for each parameter. In order to draw from the posterior distribution, we use Gibbs sampling, a MCMC method that consists on breaking the parameter vector in several blocks so that each block's posterior distribution conditional on the observed data as well as on the other blocks has a convenient form to draw values from. If we repeatedly draw values from these blocks, these values eventually converge to draws from the joint distribution of the entire parameter vector defined in (1.17) (see chapter 9 of Train (2009) for details).

In each iteration, we use Gibbs sampling to break the posterior distribution (1.17) into 3 blocks. In the first block, we draw $\bar{\theta}$ and θ_σ from their conditional posterior distributions - the Normal distribution for each term of $\bar{\theta}$ and the Inverted Gamma distribution for each term of θ_σ. In this block, standard procedures are used to obtain the draws (see chapter 12 of Train (2009) for details). In the second block, we draw individual parameters θ_i, whose conditional posterior distribution is proportional to

$$\prod_{t=1}^{T}\prod_{j=0}^{1}\left(\frac{exp(\bar{v}(a=j,x_{it};\theta_i,\theta_1))}{\sum_{k=0}^{1}exp(\bar{v}(a=k,x_{it};\theta_i,\theta_1))}\right)^{I\{a=j\}} pd(\theta_i)k(\bar{\theta},\theta_\sigma). \tag{1.18}$$

To draw from (1.18), we use the Metropolis-Hastings algorithm, which is also a MCMC method (see chapter 9 of Train (2009) for details). Finally, in the third block, we draw fixed parameters θ_1, whose conditional posterior distribution is proportional to

$$\prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=0}^{1}\left(\frac{exp(\bar{v}(a=j,x_{it};\theta_i,\theta_1))}{\sum_{k=0}^{1}exp(\bar{v}(a=k,x_{it};\theta_i,\theta_1))}\right)^{I\{a=j\}}. \tag{1.19}$$

Again, we use the Metropolis-Hastings algorithm to draw values from (1.19).

The Gibbs sampling allows us to obtain, for each parameter, a sequence of values, one per iteration, which is used to estimate the parameters and their standard deviations. Firstly, we discard the initial values of that sequence that constitute burn-in. Then, we use the remaining values and, for each parameter, we compute the mean of the sequence of values associated with it as well as the standard deviation of those values. The computed mean is the parameter estimate and the computed standard deviation is the estimate of the standard deviation of the parameter.

Note that in order to evaluate (1.18) and (1.19), we need to compute $\bar{v}(.)$, which is defined in equation (1.8). However, $\bar{v}(.)$ is not known since it depends on the (unknown) value function. Therefore, in addition to the MCMC step to draw from the posterior distribution, we need in each iteration a step for the computation of the value function, allowing for $\bar{v}(.)$ to be known.

In order to calculate the value function, we use the procedure suggested in Imai *et al.* (2009): instead of solving the Bellman equation in each iteration, we iterate it only once. By using this procedure, the estimation method solves the dynamic programming problem and simultaneously estimates the parameters.

Let us define θ_i^{*r} and θ_1^{*r} as the candidate parameters, respectively, of θ_i and θ_1 used by the Metropolis-Hastings algorithm in the MCMC step in a given iteration r. We calculate the expected future value $E_x\left[\bar{V}(x_{i,t+1};\theta) \mid a, x_{it}\right]$ as the weight average of n^* previous values functions, where the weights are defined by kernel densities of the difference between the candidate parameter in current iteration and the candidate parameter in previous iterations. The intuition here is that the value function is continuous in the parameter space, thus parameters which are closer to the current parameter have closer value functions. Therefore, candidate parameters closer to the current candidate parameter have more weight since the associated value functions are closer to the current value function.

Thus, for a given value of the state variables $x = (p, c, s)'$, we compute the expected future value in iteration r as

$$E_x^r\bar{V}(x, \theta_i^{*r}, \theta_1^{*r}) = \sum_{l=r-n^*}^{r-1} \bar{V}^l(x, \theta_i^{*l}, \theta_1^{*l}) \frac{K_h(\theta_1^{*r} - \theta_1^{*l})K_h(\theta_i^{*r} - \theta_i^{*l})}{\sum_{k=r-n^*}^{r-1} K_h(\theta_1^{*r} - \theta_1^{*k})K_h(\theta_i^{*r} - \theta_i^{*k})}, \quad (1.20)$$

where $E_x^r\bar{V}(x, \theta_i^{*r}, \theta_1^{*r})$ is the approximated expected future value in iteration r, K_h is the Gaussian kernel with bandwidth h and n^* is the number of past iterations used to approximate the expected future value.

Note that we have to compute (1.20) for given values of all the state variables. In order to reduce the computational burden, we make use of Rust (1997)'s random grid: instead of computing the expected future value for given values of the state variables, we randomly select in each iteration one value for p and c and we weight (1.20) with their transition probabilities (see Imai *et al.* (2009) and Ching *et al.* (2010) for an analysis of the conjunction of the Bayesian method proposed by Imai *et al.* (2009) with Rust (1997)'s random grid).

Let us define $x_1 = (p, c)'$ as the values of p and c in iteration r and let f_{x_1} represent the transition probabilities of x_1. Then, equation (1.20) is now replaced by

$$E_x^r \bar{V}(s, x_1, \theta_i^{*r}, \theta_1^{*r}) = \sum_{l=r-n^*}^{r-1} \bar{V}^l(s, x_1^l, \theta_i^{*l}, \theta_1^{*l}) \frac{K_h(\theta_1^{*r} - \theta_1^{*l}) K_h(\theta_i^{*r} - \theta_i^{*l}) f_{x_1}(x_1^l \mid x_1)}{\sum_{k=r-n^*}^{r-1} K_h(\theta_1^{*r} - \theta_1^{*k}) K_h(\theta_i^{*r} - \theta_i^{*k}) f_{x_1}(x_1^k \mid x_1)}. \quad (1.21)$$

By using (1.21) instead of (1.20), we only have to compute the expected future values for given values of s. We do not treat s as the other state variables since Rust (1997)'s random grid cannot be used when the transition law, given the parameters and the decision variables, is deterministic. We use the approximated expected future values obtained in (1.21) to compute $\bar{v}(.)$ defined in equation (1.8), which are then used to update the value function $\bar{V}(.)$ defined in equation (1.10).

To sum up, the Bayesian method used in this paper includes in each iteration two steps. One step employs the MCMC algorithm to draw values from the posterior distributions of the parameters. The other step allows for the solution of the dynamic programming model, using equation (1.21) to update the expected future value.

5 The Data

5.1 The Firm and the Ordering Process

Our data include information on a firm that sells alcoholic drinks and operates in the North of Portugal. The firm's products are stocked in a single store and sold to a variety of customers located in different regions. The firm sells its products to two types of clients: firms (e.g. restaurants) and final consumers. We do not have any information regarding the sales to each type of customers. Given that a significant number of customers are firms, we believe that the sales to this type of clients represents a significant share in the firm's total sales.

The firm does not produce any of its products; in fact, whenever it is considered adequate, the firm orders new deliveries from suppliers and sells those products to its clients. The ordering process is as follows: the firm has some regular suppliers who often meet the manager of the firm in order to define prices for all products. Whenever the firm needs more units of a given product or products, it contacts the respective supplier ordering a new delivery, and the supplier brings those products to the store.

In the ordering process, the firm bears a cost that involves the price of each product and the ordering cost. While we have no information on ordering costs, we consider that the ordering cost is composed of a fixed ordering cost (ζ_i) and a variable ordering cost (γ_i). The inclusion of both fixed and variable components in the ordering cost is due to the fact that the main component of the ordering cost is the transportation cost. The nature of the ordering process leads us to include in the transportation cost both fixed and variable components. Specifically, the transportation cost is higher, the higher the number of vans necessary to transport the requested products. However, when the firm orders a large amount of a given product, the supplier of the product makes a discount on the transportation cost or even offers it. Therefore, we include in the model a fixed ordering cost associated to the ordering process and a variable ordering cost to take into account the fact that the transportation cost is not independent of the quantity ordered.

5.2 The Database

The database contains weekly information on sales, prices, orders to suppliers and inventories for every product sold by the firm between January 2008 and June 2009. It is a balanced panel data with 66534 observations, with data on 853 different products during 78 weeks. The dataset includes the following information for every product and week: name of the product, wholesale and selling prices, sales, orders to suppliers and stock at the beginning of the week. Quantities are measured in number of bottles, while prices are measured in Euros.

Given these data, we define an ordering indicator: a binary variable a_{it} which is equal to one if the firm orders product i in time period t and equal to zero otherwise. Hence, we associate each positive value of orders to suppliers to $a_{it} = 1$ and no orders to $a_{it} = 0$.

We also compute the two indicator functions used in the model, namely $I\{y_{it} = s_{it} + q_{it}\}$ and $I\{t > t_{tax\,change}\}$. The term $I\{y_{it} = s_{it} + q_{it}\}$ is a stock-out indicator function which intends to capture the (negative) effect of a stock-out in profits. We compute the tax change indicator function $I\{t > t_{tax\,change}\}$ due to a change in the Portuguese tax policy during this period. Most of the alcoholic drinks are charged a VAT of 12%, but some of them are charged a higher VAT. In July 2008 (denoted in the model by $t_{tax\,change}$), the highest VAT changed from 21% to 20% and so $I\{t > t_{tax\,change}\}$ intends to capture the effects of this tax change in expected sales. The change in tax policy in July 2008 affects 1872 observations, including 36 products and 52 weeks.

Table 1.1 presents some descriptive statistics for the variables of interest. As before, q represents quantity of product ordered, y represents sales in number of bottles, s is the stock in number of bottles, c is the wholesale price and p denotes the sales price. The firm orders some products in 71.75% of the observations and the stock-out effect occurs in 0.32% of the observations. Orders, sales and stocks have a floating behavior

	Mean	Min	Max	St. Dev.	Skewness	Kurtosis	Pctil 25	Median	Pctil 75
q	122.4187	0	5977	319.7054	6.6018	81.9889	0	12	55
y	123.5191	0	5726	307.7122	6.1818	72.2997	6	14	56
s	274.5006	0	4585	457.8707	3.1427	14.8360	27	74	334
c	3.1864	0.64	18.31	1.8318	1.7141	9.3755	1.87	2.88	3.92
p	4.5628	0.92	22.75	2.6276	1.5013	6.7938	2.66	4.12	5.63
$I\{t > t_{tax\,change}\}$	0.0281	0	1	0.1654	5.7071	33.5706	0	0	0
$I\{y = s + q\}$	0.0032	0	1	0.0564	17.6307	311.8428	0	0	0
a	0.7175	0	1	0.4502	-0.9663	1.9337	0	1	1

Table 1.1: Descriptive Statistics

	Coefficient	St. Error	$z - test$	$P >\mid z - test \mid$
y	0.0009219	0.0001054	8.747	0.000
s	-0.0008462	0.0001057	-8.006	0.000
c	-0.8726403	0.4208984	-2.073	0.038
p	-0.5068481	0.3103980	-1.633	0.102

Log Likelihood = -18996.405	
LR $\chi^2(4) = 174.97$	Prob $> \chi^2 = 0.000$

Table 1.2: Fixed-Effects Logit Model for the Discrete Ordering Choice ($a = 1$ or $a = 0$)

in the sense that these variables have relatively high standard deviations. In fact, these three variables seem to have a lot of low values (the median is 12 for orders, 14 for sales and 74 for stocks) and some "peaks" explaining the difference between the mean and the median (indeed, orders, sales and stocks reach maximum values of 5977, 5726 and 4585 respectively).

Interestingly, 50% of the orders are associated with a value of 12 or an even lower number of bottles. As far as transportation costs are concerned, the low median of q may mean that a given supplier delivers in the same ordering process a small number of different products, since it is unlikely the firm is willing to bear a transportation cost just for the transportation of 12 bottles of a given product. Unfortunately, we do not have additional information to confirm that.

Table 1.2 presents a reduced form estimation of the discrete ordering decision. The explanatory variables in this model include sales, stocks, the wholesale price and the sales price. The model was estimated using the Fixed-Effects Logit estimator controlling for the existence of unobserved product heterogeneity. We note that the Likelihood Ratio (LR) test clearly points out the global significance of the model. Also, the signs of the coefficients are as expected: positive for sales and negative for stocks, the wholesale price and the sales price. Interestingly, while sales, stocks and the wholesale price are significant at a significance level of 5%, the sales price coefficient is not significant. This may be related to the existence of stocks: the firm may order the products in some periods to take advantage of discounts at the wholesale market and then keep them in stock for future sales.

6 Estimation Results

We estimate the structural model drawing from the posterior distribution of the parameters in (1.17) 30000 times. We drop the first 10000 iterations and we compute the means and standard deviations using the values from iteration 10001 to 30000. We do not estimate the discount factor δ, which is set equal to 0.985. The estimation results for the structural model are shown in Table 1.3.

The results show that the efficiency terms are statistically different from zero. Specifically, σ_{β^T} is statistically significant, suggesting there is a difference in efficiency across products. The parameter measuring losses in profits due to a stock-out is significant, although not very different across products. Interestingly, while the parameter associated with the fixed ordering cost is relevant and different across products, the parameter associated with the variable cost of ordering is neither significant nor different across products. The heterogeneity in fixed ordering costs may be explained by the fact that the firm has different suppliers for different products. In fact, although the general characteristics of the ordering process are the same for all suppliers, it is possible there are some differences in the ordering process across the suppliers, implying differences in the ordering cost. As mentioned before, we do not have detailed information on the ordering process. The decrease on taxes in July 2008 seems to have a significant impact on sales as λ has the expected sign and is statistically significant at a significance level of 5%. In fact, the decrease on taxes implied a 32% increase on sales, suggesting this type of products is very price-sensitive. As expected, the unit storage cost μ is also statistically significant.

Note that, apart from the estimates of the fixed parameters λ and μ, the values of the coefficients in column 1 of Table 1.3 are not the final estimates of the parameters. As far as γ_i, ω_i and ζ_i are concerned, we consider a lognormal distribution and we

	Coefficient	St. Error	$z-test$	$P>\mid z-test \mid$
$\overline{\beta^T}$	0.9346019	0.0677182	13.801	0.000
σ_{β^T}	0.1981808	0.0262467	7.551	0.000
$\overline{ln\,\omega}$	-4.2880810	1.6767339	-2.557	0.011
$\sigma_{ln\,\omega}$	2.0260520	1.6416476	1.234	0.217
$\overline{ln\,\gamma}$	-12.0491569	52.2364546	-0.230	0.818
$\sigma_{ln\,\gamma}$	3.5025554	4.7131788	0.743	0.457
$\overline{ln\,\zeta}$	-0.6439362	0.1558840	-4.131	0.000
$\sigma_{ln\,\zeta}$	1.2229765	0.5091166	2.402	0.016
λ	0.3205798	0.1371140	2.338	0.019
μ	0.1398554	0.0477591	2.928	0.003

Table 1.3: Estimation Results for the Structural Model ($\delta = 0.985$)

	Coefficient		Coefficient
$\overline{\beta}$	0.8156247	$\overline{\gamma}$	0.0026981
σ_β	0.1316712	σ_γ	1.2445017
$\overline{\omega}$	0.1069239	$\overline{\zeta}$	1.1094890
σ_ω	0.8257122	σ_ζ	2.0644642
λ	0.3205798	μ	0.1398554

Table 1.4: Final Values of the Means of the Parameters

parametrize it using the associated normal distribution, that is, γ_i, ω_i and ζ_i follow a lognormal distribution if and only if $ln\gamma_i \sim N(\overline{ln\,\gamma}, \sigma^2_{ln\,\gamma})$, $ln\omega_i \sim N(\overline{ln\,\omega}, \sigma^2_{ln\,\omega})$ and $ln\zeta_i \sim N(\overline{ln\,\zeta}, \sigma^2_{ln\,\zeta})$. Therefore, we have estimated the means and standard deviations of the natural logarithm of the coefficients. By using the corresponding estimates in Table 1.3, we are able to obtain the means and standard deviations of γ_i, ω_i and ζ_i (see details in appendix C).

Similarly, we have to adjust the values of the coefficients associated to β_i. Recall that we have defined that $\beta_i \sim TN_{[0,1]}(\overline{\beta^{\mathbf{T}}}, \sigma^2_{\beta^T})$. Although we consider β_i to follow a truncated normal distribution, the estimated mean and standard deviation do not take into account the fact that β_i lies between 0 and 1, that is, they do not give us the mean and standard deviation of β_i given that $0 < \beta_i \leq 1$. Therefore, we use the estimated values of $\overline{\beta^{\mathbf{T}}}$ and σ_{β^T} to compute the adjusted values for the mean and standard deviation of β_i (see details in appendix C).

Let us define $\overline{\beta}$, σ_β, $\overline{\gamma}$, σ_γ, $\overline{\omega}$, σ_ω, $\overline{\zeta}$ and σ_ζ as the adjusted values of $\overline{\beta^{\mathbf{T}}}$, σ_{β^T}, $\overline{ln\,\gamma}$, $\sigma_{ln\,\gamma}$,

$\overline{ln\,\omega}$, $\sigma_{ln\,\omega}$, $\overline{ln\,\zeta}$ and $\sigma_{ln\,\zeta}$. Table 1.4 shows the final values of the estimated parameters. The results in Table 1.4 show that the average product efficiency is around 81,6%, meaning that, on average, the firm obtains 81,6% of the maximum profit associated with a given product. Although there is some heterogeneity in efficiency across products, the estimated σ_β is not very high.

7 Counterfactual Experiments

We perform a counterfactual experiment to investigate what would be the firm's choices if its ordering decisions were fully efficient, that is, if β_i were equal to 1 for all products in all time periods. In order to do so, we simulate the dynamic programming model, using all the estimates in Table 1.3, except the estimates of the mean and standard deviation associated with β_i.

Note that in order to simulate the model, we need to compute $E\left(y_{it}\right)$. We cannot use the nonparametric estimates of $E\left(y_{it}\right)$ computed in our estimation process to avoid the Lucas' critique. For the purpose of this counterfactual experiment, we use the demand estimates in appendix D to compute $E\left(y_{it}\right)$. Also, in order to define the iid demand shock ϕ_t, we use the estimates of the residuals of the demand equation in appendix D.

We simulate the model using 100 replications. The results are displayed in Table 1.5. The first column in Table 1.5 refers to the percentage of orders in the dataset used to estimate the model, which are the actual decisions of the firm. The second column in Table 1.5 refers to the choices the firm would make in a full efficiency scenario. The results show that, in at least 14.88% of the cases, the firm decides to order products whereas the optimal decision would be no order. Thus, the results obtained with the estimation of the structural model and the counterfactual experiment indicate the existence of inefficiency in the decision process.

When the firm decides whether to order or not a given product in each time period, the sales in that period are not known. Thus, the firm bases the decision in the value of expected sales, which may be quite different from actual sales. In this dataset, the standard deviation of sales is relatively high in comparison with its mean (see Table 1.1), indicating that sales are relatively volatile. To explore the impact of volatility in sales on efficiency, we perform another counterfactual experiment. Our purpose is to reestimate the efficiency terms including in the model *stable* sales. For that, we consider the actual decisions of the firm and all the estimates in Table 1.3, except the estimates of the mean and standard deviation associated with β_i, which we intend to reestimate. We also create a new variable, *stable* sales, which is defined, for each product, as the

	Actual Decisions	Decisions in a Full Efficiency Scenario
% Orders	71.75%	56.87%
% No Orders	28.25%	43.13%

Table 1.5: Results for the Counterfactual Experiment

	Coefficient	St. Error	$z - test$	$P >\mid z - test \mid$
$\overline{\beta^T}$	0.9985082	0.0759187	13.152	0.000
σ_{β^T}	0.0712272	0.0201570	3.534	0.000

	Final Value of the Coefficient
$\overline{\beta}$	0.9426234
σ_β	0.0432069

Table 1.6: Results with *Stable* Sales

mean of the sales of that product across all time periods. In this new variable, sales are potentially different for different products but, for each product, sales are equal in all time periods. We use *stable* sales instead of actual sales to estimate the model to evaluate the impact of volatility in sales on efficiency. The results are displayed in Table 1.6.

The results show that the existence of *stable* sales leads to an increase on average efficiency from 81,6% to 94,3%, suggesting that volatility in sales is responsible for a significant part of the estimated inefficiency. We also find out that the heterogeneity in efficiency across products decreases when we consider *stable* sales, although it remains statistically significant.

8 Conclusion

In this paper, we develop a measure for dynamic (profit) efficiency in a dynamic discrete choice framework, in which decisions are over discrete rather than continuous variables. We analyze a dynamic programming inventory model and develop a measure of dynamic efficiency at the product level. For each product, we consider that, in the event of inefficiency, the firm only gets a fraction of the maximum (expected) profit for that product.

Using a dataset with weekly information on prices, sales, orders and stocks for a Portuguese firm from January 2008 to June 2009, we estimate the model with a two-stage approach. We find out that the average product efficiency is around 81,6%, implying that, on average, the firm obtains 81,6% of the maximum profit associated with a given product.

We also investigate what would be the firm's choices if its ordering decisions were fully efficient. The results show that if the firm were fully efficient, it would choose differently in at least 14.88% of the decisions. We also find out that volatility in sales is responsible for a significant part of the estimated inefficiency.

In our model, we consider that decisions over different products are separable, that is, there is no synchronization among decisions over different products so that the firm decides whether to order or not a given product on an individual basis, without taking into consideration the order decisions over other products. If there is such a synchronization effect among products, the fixed ordering cost may be shared among different products when the supplier is the same, reducing the fixed cost for any one product. So, it is possible that ignoring this synchronization effect leads to some bias in our results as the ordering behaviour for a given product may look sub-optimal: it may appear that the firm is acting non-optimally (ordering too soon or too later) while in fact it is acting rationally, trying to synchronize orders of different products to reduce the number of orders and the ordering costs. The fact that our model does not take synchronization in orders into account is due to lack of data and difficulty in implementing a model with synchronization. We do not have information on suppliers so we do not know which supplier is common to which products. Without this information, it is difficult to develop a model with synchronization among products because we do not know which are the products whose orders can or cannot be synchronized. Also, it is difficult to implement a model that takes into account synchronization in the ordering process because in this case the likelihood function is not equal to the product of the CCP for each product, as there is correlation in the ordering decisions among products.

To the best of our knowledge, this is the first paper to use a dynamic discrete choice framework to measure efficiency at a micro level. We believe that this approach is promising as it allows us to have information regarding firm efficiency that until now was only available for models with continuous decision variables. In addition, it allows us to perform counterfactual experiments and compare actual decisions with optimal decisions.

Appendix

A. Nonparametric Estimation of the Transition Probabilities of the State Variables

Following Sanchez-Mangas (2002), we generate a nonparametric estimate of $f_p(p_{i,t+1} \mid p_{it})$. The transition probability for c, $f_c(c_{i,t+1} \mid c_{it})$, is defined similarly.

Let us discretize the state variables p_{it}, c_{it} and s_{it}: we consider $M_1 = 13$ cells for p_{it}, $M_2 = 12$ cells for c_{it} and $M_3 = 13$ cells for s_{it}, so in fact we have $M = M_1 \times M_2 \times M_3 = 2028$ cells.

We denote the discretized values of the state variables by p^c_{it}, c^c_{it} and s^c_{it} and the values of the state variables in the m^{th} cell by p^m_{it}, c^m_{it} and s^m_{it}. Let us define $x^m = (p^m, c^m, s^m)$.

We estimate the transition probabilities for p as

$$\hat{Prob}(p^c_{t+1} = p^m \mid p^c_t = p^l) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} I\left\{p^c_{i,t+1} = p^m\right\} K_1\left(p_{it}, p^l\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T} K_1\left(p_{it}, p^l\right)}, \tag{1.22}$$

for $m, l = 1, ..., M_1$. K_1 is the univariate gaussian kernel defined as

$$K_1\left(p_{it}, p^l\right) = \frac{1}{(2\pi)^{1/2}} exp\left\{-\frac{1}{2}\left(\frac{p_{it} - p^l}{h_1}\right)^2\right\}, \tag{1.23}$$

where h_1 is the bandwidth parameter, defined according to the Silverman's rule.

B. Other Nonparametric Estimates

We follow Sanchez-Mangas (2002) to generate a nonparametric estimate of $E(q_{it} \mid x_{it}, a_{it} = 1)$. For $a_{it} = 0$, this term is defined by $E\left(q_{it} \mid x_{it}, a_{it} = 0\right) = 0$. The nonparametric estimates of $E\left(q_{it} \mid x_{it}, a_{it}\right)$, $E\left(I\{y_{it} = s_{it} + q_{it}\} \mid x_{it}, a_{it}\right)$, $E\left(I\{t > t_{tax\, change}\} \mid x_{it}, a_{it}\right)$ and $E\left(y_{it} \mid x_{it}, a_{it}\right)$ are defined similarly to $E\left(q_{it} \mid x_{it}, a_{it} = 1\right)$.

We start by discretizing the variable $\{q_{it}; a_{it} = 1\}$, that is, we consider the variable q_{it} only for those observations in which $q_{it} > 0$. We use a uniform grid with H cells.

We denote the value of this discretized variable by q^c and the value of the variable in cell h, $h = 1, ..., H$, by q^h. We estimate $E\left(q \mid x^m, a = 1\right)$, for $m = 1, ..., M$ as

$$\sum_{h=1}^{H} q^h Prob(q^h \mid x^m, a = 1), \tag{1.24}$$

where

$$Prob(q^h \mid x^m, a = 1) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} I\left\{q^c_{it} = q^h\right\} I\left\{a_{it} = 1\right\} K_3(x_{it}, x^m)}{\sum_{i=1}^{N}\sum_{t=1}^{T} I\left\{a_{it} = 1\right\} K_3(x_{it}, x^m)}, \tag{1.25}$$

for $h = 1, ..., H$ and $m = 1, ..., M$, where

$$K_3(x_{it}, x^m) = \frac{1}{(2\pi)^{3/2}} exp\left\{ -\frac{1}{2}\left[\left(\frac{p_{it} - p^m}{h_1}\right)^2 + \left(\frac{c_{it} - c^m}{h_2}\right)^2 + \left(\frac{s_{it} - s^m}{h_3}\right)^2 \right] \right\} \tag{1.26}$$

is the trivariate gaussian kernel and h_1, h_2 and h_3 are defined according to the Silverman's rule.

C. Computation of the Final Estimates of the Parameters

Here we show how to compute the final estimates of the means and standard deviations of the random coefficients from the estimated values in Table 1.3.

For all the elements of $\bar{\theta}$ and θ_σ, with the exception of $\overline{\beta^T}$ and σ_{β^T}, let us define a given value of $\bar{\theta}$ and its corresponding standard deviation in θ_σ by $\bar{\Upsilon}$ and σ_Υ and their adjusted values by $\bar{\Upsilon}^c$ and σ^c_Υ. Then, the final mean and the final standard deviation of the corresponding parameter is given by (see chapter 6 of Train (2009))

$$\bar{\Upsilon}^c = exp(\bar{\Upsilon} + \sigma^2_\Upsilon/2) \tag{1.27}$$

and

$$\sigma^c_\Upsilon = \sqrt{exp(2\bar{\Upsilon} + \sigma^2_\Upsilon) \times (exp(\sigma^2_\Upsilon) - 1)}. \tag{1.28}$$

For $\overline{\beta^T}$ and σ_{β^T}, let us define the adjusted values by $\overline{\beta}$ and σ_β respectively. We can compute such values as (see Johnson *et al.* (1994))

$$\overline{\beta} = \overline{\beta^T} + \frac{\varphi\left(\frac{-\overline{\beta^T}}{\sigma_{\beta^T}}\right) - \varphi\left(\frac{1-\overline{\beta^T}}{\sigma_{\beta^T}}\right)}{\Phi\left(\frac{1-\overline{\beta^T}}{\sigma_{\beta^T}}\right) - \Phi\left(\frac{-\overline{\beta^T}}{\sigma_{\beta^T}}\right)} \times \sigma_{\beta^T} \tag{1.29}$$

and

$$\sigma_\beta = \sqrt{\sigma^2_{\beta^T} \times \left[1 + \frac{\frac{-\overline{\beta^T}}{\sigma_{\beta^T}}\varphi\left(\frac{-\overline{\beta^T}}{\sigma_{\beta^T}}\right) - \frac{1-\overline{\beta^T}}{\sigma_{\beta^T}}\varphi\left(\frac{1-\overline{\beta^T}}{\sigma_{\beta^T}}\right)}{\Phi\left(\frac{1-\overline{\beta^T}}{\sigma_{\beta^T}}\right) - \Phi\left(\frac{-\overline{\beta^T}}{\sigma_{\beta^T}}\right)} - \left(\frac{\varphi\left(\frac{-\overline{\beta^T}}{\sigma_{\beta^T}}\right) - \varphi\left(\frac{1-\overline{\beta^T}}{\sigma_{\beta^T}}\right)}{\Phi\left(\frac{1-\overline{\beta^T}}{\sigma_{\beta^T}}\right) - \Phi\left(\frac{-\overline{\beta^T}}{\sigma_{\beta^T}}\right)}\right)^2\right]}, \tag{1.30}$$

where φ and Φ denote the standard normal probability density function and the standard normal cumulative distribution function respectively.

D. Estimation of the Demand Parameters

For those observations where $y_{it} < s_{it} + q_{it}$, that is, when there are no stockouts, we follow Aguirregabiria (1999) and estimate the demand

$$log\, y_{it} = \eta_i^0 + \eta\, log\, p_{it} + \phi_{it}. \tag{1.31}$$

The estimation of equation (1.31) using standard methods poses two problems. First, there are brand fixed-effects. Second, prices may be correlated with the random component ϕ_{it} (see chapter 13 of Train (2009) for details) and so there is endogeneity.

In order to take into account the existence of brand-specific effects, we use first differences and estimate equation (1.31) as

$$\triangle log\, y_{it} = \eta \triangle log\, p_{it} + \triangle \phi_{it}. \tag{1.32}$$

We estimate equation (1.32) using Instrumental Variables (IV) in order to take into account the endogeneity in prices. We use the Two-Stages Least Squares (2SLS) estimator in which $\triangle log\, p_{it}$ is instrumented by $\triangle log\, p_{it-2}$ and $\triangle log\, p_{it-3}$. The results are shown in Table 1.7.

The demand coefficient η has the expected sign and it is statistically significant at the usual significance levels. The Sargan test for overidentifying restrictions shows that the null hypothesis is not rejected, pointing out the consistency of the instruments used.

	Coefficient	St. Error	$t-test$	$P >\mid t-test \mid$
η	-2.152188	0.330468	-6.51	0.000

Sargan test $\chi^2(1) = 2.616130$	Prob $> \chi^2 = 0.1058$

Table 1.7: Results for the IV Estimation

References

[1] Aguirregabiria, V. (1999), “The Dynamics of Markups and Inventories in Retailing Firms”, Review of Economic Studies, 66, 275-308.

[2] Aguirregabiria, V. and P. Mira (2010), “Dynamic Discrete Choice Structural Models: A Survey”, Journal of Econometrics, 156, 38-67.

[3] Arrow, K., T. Harris, and J. Marschak (1951), “Optimal Inventory Policy”, Econometrica, 19, 250-272.

[4] Berger, A. and L. Mester (1997), "Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?", Journal of Banking & Finance, 21, 895-947.

[5] Blinder, A. (1981), "Retail Inventory Investment and Business Fluctuations", Brookings Papers on Economic Activity, 2, 443-505.

[6] Chambers, R. G., Chung, Y. and R. Färe (1998), "Profit, Directional Distance Functions, and Nerlovian Efficiency", Journal of Optimization Theory and Applications, 98, 351-364.

[7] Ching, A., S. Imai, M. Ishihara, and N. Jain (2010), "A Guide to Bayesian Estimation of Dynamic Discrete Choice Models with an Application to a Store-Level Reward Program", Available at SSRN: http://ssrn.com/abstract=1398444.

[8] Deneckere, R., H. Marvel, and J. Peck (1996), "Demand Uncertainty, Inventories and Resale Price Maintenance", The Quarterly Journal of Economics, 111, 885-913.

[9] Färe, R., S. Grosskopf, and W. Weber (2004), "The Effect of Risk-Based Capital Requirements on Profit Efficiency in Banking", Applied Economics, 36, 1731-1743.

[10] Hall, G. and J. Rust (2000), "An Empirical Model of Inventory Investment by Durable Commodity Intermediaries", Carnegie-Rochester Conference Series on Public Policy, 52, 171-214.

[11] Imai, S., N. Jain, and A. Ching (2009), "Bayesian Estimation of Dynamic Discrete Choice Models", Econometrica, 77, 1865-1899.

[12] Johnson, N., S. Kotz, and N. Balahrishnan (1994), "Continuous Univariate Distributions", New York, 2nd edition.

[13] Kumbhakar, S. and K. Lovell (2000), "Stochastic Frontier Analysis", Cambridge University Press, New York.

[14] Lasserre, P. and P. Ouellette (1999), "Dynamic Factor Demands and Technology Measurement under Arbitrary Expectations", Journal of Productivity Analysis, 11, 219-241.

[15] Magnac, T. and D. Thesmar (2002), "Identifying Dynamic Discrete Decision Processes", Econometrica, 70, 801-816.

[16] Nemoto, J. and M. Goto (1999), “Dynamic Data Envelopment Analysis: Modeling Intertemporal Behavior of a Firm in the Presence of Productive Inefficiencies”, Economic Letters, 64, 51-56.

[17] Nemoto, J. and M. Goto (2003), “Measurement of Dynamic Efficiency in Production: An Application of Data Envelopment Analysis to Japanese Electric Utilities”, Journal of Productivity Analysis, 19, 191-210.

[18] Ouellette, P. and L. Yan (2008), “Investment and Dynamic DEA”, Journal of Productivity Analysis, 29, 235-247.

[19] Ramey, V. and K. West (1999), “Inventories”, in J. Taylor and M. Woodford (Eds.), Handbook of Macroeconomics, 1B, North-Holland, Amsterdam, 863-923.

[20] Rust, J. (1994), “Structural Estimation of Markov Decision Processes”, in R. E. Engle and D. McFadden (Eds.), Handbook of Econometrics, 4, North-Holland, Amsterdam, 3081-3143.

[21] Rust, J. (1997), “Using Randomization to Break the Curse of Dimensionality”, Econometrica, 65, 487-516.

[22] Sanchez-Mangas, R. (2002), “Pseudo-Maximum Likelihood Estimation of a Dynamic Structural Investment Model”, Working Paper 02-62, Statistics and Econometrics Series, Universidad Carlos III de Madrid.

[23] Scarf, H. (1959), “The Optimality of (S; s) Policies in the Dynamic Inventory Problem”, in K. Arrow, S. Karlin and P. Suppes (Eds.), Mathematical Methods for the Social Sciences, Stanford University Press, Stanford, 196-202.

[24] Serra, T., A. O. Lansink and S. Stefanou (2011), “Measurement of Dynamic Efficiency, a Directional Distance Function Parametric Approach”, American Journal of Agricultural Economics, 93, 756-767.

[25] Silva, E. and S. Stefanou (2007), “Dynamic Efficiency Measurement: Theory and Application”, American Journal of Agricultural Economics, 89, 398-419.

[26] Train, K. (2009), “Discrete Choice Methods with Simulation”, Cambridge University Press, 2nd Edition.

Essay 2: Dynamic Efficiency and Machine Replacement: A Discrete Choice Approach

Abstract

We use a dynamic discrete choice framework to analyze efficiency at the firm level. We consider a dynamic programming machine replacement model which allows for the existence of firm heterogeneity as well as efficiency heterogeneity across firms. We estimate the structural parameters using the Bayesian estimation method proposed by Imai *et al.* (2009), which easily accommodates the existence of random parameters. Our counterfactual experiments suggest that different types of firms should react differently in order to be fully efficient: big firms should do more replacements and small firms should replace less.

Keywords: Dynamic Machine Replacement Model, Dynamic Efficiency, Bayesian estimation Methods

JEL Classification: C15, C25, C61, D21

1 Introduction

Dynamic efficiency at the firm level has been evaluated within models in which firms decide over continuous variables (e.g., see Silva and Stefanou (2007), Nemoto and Goto (1999, 2003), Lasserre and Ouellette (1999), Ouellette and Yan (2008) and Serra *et al.* (2011)). For instance, Silva and Stefanou (2007) considers an adjustment-cost model of the firm - where the firm's choices concern the amount of variable and dynamic factors in each time period - and proposes a measure of dynamic cost efficiency within this framework. However, there are many situations in which firms do not make decisions over continuous variables, but rather over discrete variables. Examples include decisions on product ordering, patent renewal, machine replacement, price changes, among others.

The theoretical framework based on continuous choices may not be an adequate representation of firm's behaviour when discrete choice decisions are considered. In fact, these decision processes are more properly analyzed using the framework based on the discrete choice literature. Since the seminal work of McFadden (1978) and its extension to Markov decision problems by Rust (1987), Hotz and Miller (1993) and Aguirregabiria and Mira (2002), among others, the discrete choice framework has been used in a wide variety of applications in which agents decide over discrete variables (e.g., see Aguirregabiria and Mira (2010) for a survey on applications of discrete choice models in a dynamic environment).

In a machine replacement problem, firms have to decide, in each period, whether to replace their machines or not so that their intertemporal costs are minimized. This problem has been analyzed using the dynamic discrete choice framework (e.g., see Rust (1987), Das (1992), Kennet (1994), Miranda and Schnitkey (1995), Rust and Rothwell (1995) and Kasahara (2009)). In these models, the principle of revealed preferences and data on firms' choices and outcomes are used to estimate the structural parameters. These parameters are structural in the sense that they have precise economic meanings, representing agent's payoffs, preferences and beliefs about future events. One of the advantages of this approach is the possibility of performing counterfactual experiments, that is, to evaluate the impact of changes in structural parameters or exogenous variables on firms' decisions.

In this paper, we use a dynamic discrete choice framework to analyze efficiency at the firm level. Like some of the papers on dynamic efficiency mentioned above, we consider that firms attempt to minimize intertemporal costs. In contrast with the other studies, we explicitly assume that firms make decisions over discrete variables. Specifically, we analyze a machine replacement problem and introduce a measure of dynamic efficiency in this model. Inefficiency arises whenever firms deviate from their optimal, forward-looking, rational choices, therefore getting a cost greater than the minimum cost.

In our model, the structural parameters are random parameters to allow for the existence of firm heterogeneity as well as efficiency heterogeneity across firms. We use a dataset with yearly information for 290 Portuguese manufacturing firms from 2001 to 2008 and estimate the model for all the firms in our dataset, as well as for subsamples on big and small firms. The model is estimated with a two-stage approach. In the first stage, we estimate the transition probabilities of the state variables using nonparametric methods. In the second stage, we use the method proposed by Imai *et al.* (2009), which is a Bayesian estimation method that allows simultaneously for the solution of the dynamic programming problem and the estimation of the parameters. This method includes two steps in each iteration: one step solves the dynamic programming model and the other one employs the Markov Chain Monte Carlo (MCMC) algorithm to draw values from the posterior distributions of the parameters. We also perform counterfactual experiments to know what would be the firms' optimal choices and compare them with the actual choices. The counterfactuals results suggest that big firms should have done more replacements and small firms should have replaced less in order to be fully efficient.

The paper is organized as follows. In Section 2, we present the machine replacement model with inefficiency. We discuss the estimation method in Section 3. The data and the estimation results are reported in Sections 4 and 5, respectively. Our counterfactual experiments are discussed in Section 6. Section 7 concludes the paper.

2 The Machine Replacement Model with Inefficiency

We present a discrete choice dynamic programming model where each firm decides, in each period, whether to replace its machine or not. Time is discrete and indexed by t, $t = 0, ..., \infty$, while firms are indexed by i, i = 1,..., N. The decision variables, a_{it}, belong to a discrete, finite set $A = \{0, 1\}$, where $a_{it} = 1$ if firm i replaces its machine in time period t and $a_{it} = 0$ otherwise. The current cost function for firm i is given by

$$C(a_{it}, s_{it}) = a_{it}(c_{it} + \gamma_i) + \omega_i x_{it}, \tag{2.1}$$

where c_{it} represents the net replacement cost of the machine (i.e., the price of the new machine minus the scrap value of the machine), x_{it} is the age of the machine and s_{it} represents the observed state variables, $s_{it} = (c_{it}, x_{it})'$.

The current cost function specified in (2.1) formalizes the trade-off associated with the machine replacement decision: a firm which decides to keep its machine saves the replacement cost, yet older machines have higher maintenance costs. Therefore, there is a trade-off between minimizing replacement costs and minimizing maintenance costs.

The net replacement cost follows an exogenous first-order markov process $f_c(c_{it+1} \mid c_{it})$. The age of the machine evolves as follows: $x_{it+1} = (1 - a_{it})x_{it} + 1$.

Parameters in the current cost function include γ_i and ω_i. The structural parameter γ_i is included in total replacement costs and represents all the remaining costs (other than the net replacement cost c_{it}) that a firm bears when it decides to replace its machine. For instance, γ_i may represent a learning cost associated with the utilization of a new machine. The component $\omega_i x_{it}$ represents age-specific maintenance costs.

Both parameters included in the current cost function are random parameters to take into account firm heterogeneity. We assume a lognormal distribution for γ_i, $ln\gamma_i \sim N(\overline{ln\,\gamma}, \sigma^2_{ln\,\gamma})$ and a normal distribution for ω_i, $\omega_i \sim N(\overline{\omega}, \sigma^2_{\omega})$.

The current cost function for firm i in (2.1) represents firm i's minimum cost. Allowing for inefficiency in the replacement decisions, the current cost function in (2.1) is redefined as

$$C(a_{it}, s_{it}) = \beta_i \times [a_{it}(c_{it} + \gamma_i) + \omega_i x_{it}]\,,$$

with $\beta_i \geq 1$. The parameter β_i is the firm-specific efficiency measure. If firm i's replacement choice is optimal, then $\beta_i = 1$; if firm i's replacement choice is not optimal, $\beta_i > 1$ and the firm gets a cost which is greater than the minimum cost.[1]

We allow for efficiency heterogeneity across firms by treating β_i as a random parameter. Specifically, we assume a truncated normal distribution for β_i: $\beta_i \sim TN_{[1,\infty[}(\overline{\beta^{\mathbf{T}}}, \sigma^2_{\beta^T})$, where $TN_{[1,\infty[}(.,.)$ is the one-sided truncated normal density with lower bound 1. Hereafter, $\theta_i = (\beta_i, ln\,\gamma_i, \omega_i)$ represents all the parameters in the current cost function.

Given the state variables, the problem of a firm is to make decisions a_{it} in order to maximize the expected discounted flow of utility over time

$$E\left(\sum_{t=0}^{\infty} \delta^t\, u(a_{it}, s_{it}, \varepsilon_{it})\right),$$

where $\delta \in (0,1)$ is the discount factor and $u(a_{it}, s_{it}, \varepsilon_{it})$ is the current utility function, $u(a_{it}, s_{it}, \varepsilon_{it}) = -C(a_{it}, s_{it}) + \varepsilon_{it}(a_{it})$. $\varepsilon_{it}(a_{it})$ is a random component, representing unobservable variables to the econometrician. We assume that ε_{it} is independent over time with type 1 extreme value distribution $G(\varepsilon_{it})$.

Let $V(s_{it}, \varepsilon_{it}; \theta_i)$ denote the value function of the dynamic programming problem. The Bellman equation is defined as

$$V(s_{it}, \varepsilon_{it}; \theta_i) = \max_{a \in A} \left\{u(a, s_{it}, \varepsilon_{it}) + \delta E_{s,\varepsilon}\left[V(s_{it+1}, \varepsilon_{it+1}; \theta_i) \mid a, s_{it}, \varepsilon_{it}\right]\right\}. \tag{2.2}$$

We can rewrite problem (2.2) using the concept of integrated Bellman equation (see Aguirregabiria and Mira (2010))

$$\begin{aligned}\bar{V}(s_{it}; \theta_i) &\equiv \int V(s_{it}, \varepsilon_{it}; \theta_i)\, dG(\varepsilon_{it}) \\ &= \int \max_{a \in A} \left\{u(a, s_{it}, \varepsilon_{it}) + \delta E_s\left[\bar{V}(s_{i,t+1}; \theta_i) \mid a, s_{it}\right]\right\} dG(\varepsilon_{it})\end{aligned}$$

[1] We have also considered the possibility of introducing an additive component in the current cost function to measure efficiency. However, in discrete choice models “only differences in utility matter” (see Train (2009)) and such a term would not be identifiable.

$$= log \left(\sum_{a \in A} exp \left\{ \bar{v}(a, s_{it}; \theta_i) \right\} \right), \tag{2.3}$$

where the last equality results from the assumption that ε_{it} follows a type 1 extreme value distribution and

$$\bar{v}(a, s_{it}; \theta_i) \equiv -C(a, s_{it}) + \delta E_s \left[\bar{V}(s_{i,t+1}; \theta_i) \mid a, s_{it} \right]. \tag{2.4}$$

Denote the optimal rule by $\bar{\alpha}(s_{it}, \varepsilon_{it}; \theta_i) = arg\, max_{a \in A} \left\{ \bar{v}(a, s_{it}; \theta_i) + \varepsilon_{it}(a) \right\}$. The Conditional Choice Probability (CCP), which is a component of the likelihood function, is defined as

$$P(a \mid s; \theta_i) = \int I \left\{ \bar{\alpha}(s, \varepsilon; \theta_i) = a \right\} dG(\varepsilon)$$

$$= \int I \left\{ \bar{v}(a, s_{it}; \theta_i) + \varepsilon_{it}(a) > \bar{v}(a', s_{it}; \theta_i) + \varepsilon_{it}(a') \, for\, all\, a' \right\} dG(\varepsilon_{it})$$

$$= \frac{exp \left\{ \bar{v}(a, s_{it}; \theta_i) \right\}}{\sum_{j=0}^{1} exp \left\{ \bar{v}(a = j, s_{it}; \theta_i) \right\}}, \tag{2.5}$$

where the last equality follows from the assumption that ε_{it} follows a type 1 extreme value distribution.

Having data on $i = 1, ..., N$ firms during $t = 1, ..., T_i$ periods, where T_i is the number of periods firm i is observed, we can define the (conditional) likelihood function as

$$L(\theta_i) = \prod_{i=1}^{N} \prod_{t=1}^{T_i} \prod_{j=0}^{1} P(a \mid s; \theta_i)^{I\{a=j\}}. \tag{2.6}$$

3 Estimation Method

We estimate the model using a two-stage approach. In the first stage, we estimate the transition probability of c_{it} using nonparametric methods (see appendix A for details). In the second stage, we estimate the structural parameters using the Bayesian method developed in Imai *et al.* (2009) and also analyzed in Ching *et al.* (2010). This method employs Markov Chain Monte Carlo (MCMC) algorithms and Bayesian methods, which consist on specifying priors and proposal distributions for the parameters and then drawing many values from the posterior distribution of the parameters conditional on the observed data.

Denoting all parameters in the current cost function by θ_i, $\theta_i = (\beta_i, ln\,\gamma_i, \omega_i)$, set $\bar{\theta} = (\overline{\beta^T}, \overline{ln\,\gamma}, \overline{\omega})$ and $\theta_\sigma = (\sigma_{\beta^T}, \sigma_{ln\,\gamma}, \sigma_\omega)$. We consider a Normal prior for each term of $\bar{\theta}$ and an Inverted Gamma prior for each term of θ_σ. The proposal distributions for θ_i have already been specified in Section 2. Our purpose is, therefore, to estimate the parameters in $\bar{\theta}$ and θ_σ. We do not estimate the discount factor, δ, because in this type of models δ is nonparametrically non-identified (see Rust (1994), Magnac and Thesmar (2002) for details). We assume that δ is known and equal to 0.975.

The posterior distribution of the parameters, $\Lambda(.)$, is defined as

$$\Lambda(\theta_i, \bar{\theta}, \theta_\sigma) \propto L(\theta_i) pd(\theta_i) k(\bar{\theta}, \theta_\sigma), \tag{2.7}$$

being proportional to a function that depends on the likelihood function $L(\theta_i)$ defined in (2.6), the priors $k(\bar{\theta}, \theta_\sigma)$ and the proposal distributions $pd(\theta_i)$.

We use Gibbs sampling to draw values from the posterior distribution in (2.7). This MCMC method allows us to break the parameter vector in several blocks so that each block's posterior distribution conditional on the observed data and on the other blocks has a convenient form to draw values from (see chapter 9 of Train (2009) for details). We break the posterior distribution in (2.7) into 2 blocks. In the first block, we draw $\bar{\theta}$ and θ_σ from their conditional posterior distributions - the Normal distribution for each term of $\bar{\theta}$ and the Inverted Gamma distribution for each term of θ_σ. In this block, we make use of standard procedures to obtain the draws (see chapter 12 of Train (2009) for details). In the second block, we draw individual parameters θ_i, whose conditional posterior distribution is proportional to

$$\prod_{t=1}^{T_i} \prod_{j=0}^{1} \left(\frac{exp(\bar{v}(a=j, s_{it}; \theta_i))}{\sum_{k=0}^{1} exp(\bar{v}(a=k, s_{it}; \theta_i))} \right)^{I\{a=j\}} pd(\theta_i) k(\bar{\theta}, \theta_\sigma). \tag{2.8}$$

To draw values from (2.8), we use the Metropolis-Hastings algorithm (see chapter 9 of Train (2009) for details).

With this procedure, we obtain, for each parameter, a set of values, one per iteration, which we use to estimate the parameters and their standard deviations. We discard the initial values that constitute burn-in and compute, for each parameter, the average and the standard error of the remaining values. This average and this standard error correspond to the parameter estimate and the estimate of its standard deviation.

In order to draw from (2.8), we need to know $\bar{v}(.)$, which is defined in (2.4). However,

$\bar{v}(.)$ is not known since it depends on the (unknown) value function. We follow Imai *et al.* (2009) to compute the value function: we iterate the Bellman equation only once, instead of solving the dynamic programming problem in each iteration. Let θ_i^{*r} denote the candidate parameter of θ_i in a given iteration r, which is used in the Metropolis-Hastings algorithm within the MCMC step. The expected future value $E_s\left[\bar{V}(s_{i,t+1};\theta_i) \mid a, s_{it}\right]$ is defined as the weight average of n^* previous values functions. For a given value of the state variables, we compute the expected future value in iteration r as

$$E_s^r\bar{V}(s,\theta_i^{*r}) = \sum_{l=r-n^*}^{r-1} \bar{V}^l(s,\theta_i^{*l})\frac{K_h(\theta_i^{*r}-\theta_i^{*l})}{\sum_{k=r-n^*}^{r-1} K_h(\theta_i^{*r}-\theta_i^{*k})}, \tag{2.9}$$

where $E_s^r\bar{V}(s,\theta_i^{*r})$ is the approximated expected future value in iteration r, K_h is the Gaussian kernel with bandwidth h and n^* is the number of past iterations used to approximate the expected future value.

The approximated expected future values defined in (2.9) allow us to compute $\bar{v}(.)$ defined in (2.4), which are then used to update the value function $\bar{V}(.)$ defined in (2.3).

In sum, the estimation method includes, in each iteration, two steps. In one step, we use the MCMC algorithm to draw values from the posterior distributions of the parameters. In the other step, we use equation (2.9) to update the expected future value, allowing for the solution of the dynamic programming model.

4 The Data

The data is taken from Bureau Van Dijk's SABI database. The database includes yearly information on 290 Portuguese manufacturing firms from the textile and clothing industries between 2001 and 2008. It is an unbalanced panel with 2203 observations and 4 to 8 years of information for each firm. Table 2.1 presents more details about the number of firms and the number of observations.

The variables we use to estimate the model include the net replacement cost c_{it}, the age of the machine in years, x_{it}, a dummy variable, D_i, indicating whether the firm is a small firm or a big firm, and the replacement decision a_{it}.

With respect to the net replacement cost c_{it}, we use values for machinery in Millions of Euros. Following Kasahara (2009), we assume that the scrap value is equal to zero and so c_{it} is equal to the cost of the new machine.

The Dummy variable D_i is equal to 1 if firm i is a big firm and is equal to zero otherwise. We include D_i in our dataset to investigate whether the results for small

Total Number of Observations	2203
Total Number of Firms	290
Number of Firms with 4 Years of Data	1
Number of Firms with 5 Years of Data	6
Number of Firms with 6 Years of Data	28
Number of Firms with 7 Years of Data	39
Number of Firms with 8 Years of Data	216

Table 2.1: Number of Observations and Firms

and big firms are very different or are qualitatively the same. We use firms' sales to define whether the firm is a small or a big firm. A firm is considered to be a big firm if its sales are greater than the average sales across firms; otherwise, we set $D_i = 0$ and the firm is considered to be a small firm.

As far as the control variable is concerned, we follow Cooper *et al.* (1999) and Kasahara (2009) and use the investment rate in equipment to define a_{it}. If the investment rate is greater than 20%, then a_{it} is equal to 1; otherwise, $a_{it} = 0$. Here, the investment rate is defined as the value of equipment in current year over the corresponding value in the previous year minus 1. For the purpose of the computation of the investment rate, we only use values relative to machinery and equipment (i.e., values relative to, for instance, buildings and vehicles are excluded).

Table 2.2 presents some descriptive statistics. In the data, big firms represent 24.15% of the total number of firms. Machine replacement occurs in 30% of the observations. The percentage of replacement is similar for both types of firms: 31.02% for big firms and 29.68% for small firms. On average, the age of the machine is around 2.3 years and very similar for both types of firms. This mean value is somewhat above the median value of 2, which can be explained by the existence of older machines (the maximum value for the age of the machine is 7 years). The net replacement cost is, on average, 1.1026 Millions of Euros and, for a significant part of the observations, it lies between 0.222 and 1.336 Millions of Euros (these values are, respectively, the 25% and 75% percentiles). There are, however, some very low and high values: the maximum value in our data is around 12.16 Millions of Euros and the minimum value is around 0.001 Millions of Euros. The existence of these very low and high values may explain the difference between the median and the mean of the variable. Table 2.2 also shows that big firms bear higher net replacement costs than small firms: the average value and the standard deviation of the net replacement cost are higher for big firms than for small

All Firms	Mean	Min	Max	St. Dev.	Skewness	Kurtosis	Pctil 25	Median	Pctil 75
c	1.1026	0.001	12.16	1.5283	2.9362	13.9702	0.222	0.533	1.336
x	2.2637	1	7	1.3948	1.0639	3.4880	1	2	3
a	0.3000	0	1	0.4584	0.8726	1.7615	0	0	1
D	0.2415	0	1	0.4281	1.2080	2.4594	0	0	0

Big Firms	Mean	Min	Max	St. Dev.	Skewness	Kurtosis	Pctil 25	Median	Pctil 75
c	2.2314	0.004	12.16	2.2041	1.8599	6.6288	0.739	1.627	2.809
x	2.2726	1	7	1.4218	1.0863	3.5463	1	2	3
a	0.3102	0	1	0.4630	0.8209	1.6738	0	0	1

Small Firms	Mean	Min	Max	St. Dev.	Skewness	Kurtosis	Pctil 25	Median	Pctil 75
c	0.7432	0.001	7.344	1.0000	2.9277	13.3359	0.174	0.401	0.850
x	2.2609	1	7	1.3865	1.0557	3.4644	1	2	3
a	0.2968	0	1	0.4570	0.8894	1.7911	0	0	1

Table 2.2: Descriptive Statistics

firms.

5 Estimation Results

We estimate the model drawing from the posterior distribution of the parameters in (2.7) 20000 times. We do not consider the first 10000 iterations, which constitute burn-in, and we compute the means and standard deviations of the parameters using the values from iteration 10001 to 20000. As previously stated, we do not estimate the discount factor and present results for the dynamic model (with $\delta = 0.975$) and the static model (i.e., $\delta = 0$).

Table 2.3 presents the estimation results using the entire dataset (2203 observations) for both the dynamic and static models. Note that, as far as β_i and γ_i are concerned, the estimates of $\overline{\beta^T}$, σ_{β^T}, $\overline{ln\,\gamma}$ and $\sigma_{ln\,\gamma}$ are not our estimates of interest. As for β_i, we have defined that $\beta_i \sim TN_{[1,\infty[}(\overline{\beta^{\mathbf{T}}}, \sigma^2_{\beta^T})$. Despite considering that β_i follows a truncated normal distribution, the estimated mean and standard deviation $\overline{\beta^T}$ and σ_{β^T} do not take into account the fact that β_i is not less than 1. Thus, $\overline{\beta^T}$ and σ_{β^T} are not the mean and standard deviation of β_i given that $\beta_i \geq 1$. Let us define $\overline{\beta}$ and σ_β as the final, corrected mean and standard deviation of β_i. We use the estimated

	$\delta = 0.975$			$\delta = 0$		
	Coefficient	St. Error	$P >\mid z - test \mid$	Coefficient	St. Error	$P >\mid z - test \mid$
$\overline{\beta^T}$	1.3022171	0.0329280	0.000	1.2132087	0.0137768	0.000
σ_{β^T}	0.1646140	0.0180936	0.000	0.1349933	0.0070161	0.000
$\overline{ln\,\gamma}$	-10.1531674	5.4048566	0.060	-22.9578313	12.4532859	0.065
$\sigma_{ln\,\gamma}$	2.4437609	1.0632920	0.022	1.9298521	1.7318316	0.265
$\overline{\omega}$	0.0952807	0.0306173	0.002	7.5989280	7.1647563	0.289
σ_{ω}	0.2940043	0.0363423	0.000	1.5203975	1.2255828	0.215
$\overline{\beta}$	1.3148099			1.2296180		
σ_{β}	0.1520968			0.1202303		
$\overline{\gamma}$	0.0007715			0.0000000		
σ_{γ}	0.0152607			0.0000000		
Obs.	2203			2203		
Log-L	-1583.524755			-1656.033469		
LR test	$\chi^2(1) = 145.017428$					
	$P > \chi^2(1) = 0.000$					

Table 2.3: Estimation Results - All Firms

values of $\overline{\beta^{\mathbf{T}}}$ and σ_{β^T} to obtain the estimates of $\overline{\beta}$ and σ_{β}, which are reported in Table 2.3 (see details in appendix B). As far as γ_i is concerned, we have considered a lognormal distribution and we parametrize it using the associated normal distribution, that is, $ln\gamma_i \sim N(\overline{ln\,\gamma}, \sigma^2_{ln\,\gamma})$. Consequently, we have estimated the mean and standard deviation of the natural logarithm of the coefficients. Let us define the actual mean and standard deviation of γ_i as $\overline{\gamma}$ and σ_{γ}. In order to obtain these final, corrected mean and standard deviations of γ_i, we use the estimated values $\overline{ln\,\gamma}$ and $\sigma_{ln\,\gamma}$ (see appendix B for details).

The results in Table 2.3 for the dynamic model show that average efficiency across firms is around 1.31, implying that, on average, firms' actual costs are 1.31 times their minimum costs. The estimated σ_{β} is around 0.15, suggesting that there is efficiency heterogeneity across firms. This is not surprising as the data includes different types of firms - big and small firms - with potentially different characteristics. The estimated $\overline{\gamma}$ suggests that when firms decide to replace their machines, firms bear a cost (other than the replacement cost) which, on average, is around 771.5 Euros. Also, the maintenance

	$\delta = 0.975$			$\delta = 0$		
	Coefficient	St. Error	$P >\mid z - test \mid$	Coefficient	St. Error	$P >\mid z - test \mid$
$\overline{\beta^T}$	1.3613247	0.0596631	0.000	1.1877626	0.0308477	0.000
σ_{β^T}	0.2630838	0.0339326	0.000	0.1890553	0.0209829	0.000
$\overline{ln\,\gamma}$	-31.6232422	3.7905380	0.000	-90.9757056	24.0383083	0.000
$\sigma_{ln\,\gamma}$	3.0336787	1.2912939	0.019	3.3622858	6.8257226	0.622
$\overline{\omega}$	0.6793233	0.1235447	0.000	625.664465	556.7009328	0.261
σ_ω	0.7029463	0.1086410	0.000	24.3877635	81.2569942	0.764
$\overline{\beta}$	1.4059820			1.2426149		
σ_β	0.2260156			0.1497796		
$\overline{\gamma}$	0.0000000			0.0000000		
σ_γ	0.0000000			0.0000000		
Obs.		532			532	
Log-L		-518.090161			-534.013869	
LR test	$\chi^2(1) = 31.847415$					
	$P > \chi^2(1) = 0.000$					

Table 2.4: Estimation Results - Big firms

cost parameter is, on average, around 95281 Euros, which is statistically different from zero and significantly different across firms. While this average value is relatively high, it is much lower than the average replacement cost of the machine (see Table 2.2).

Interestingly, some parameter estimates of the dynamic and static models are quite different. In fact, while the estimates of $\sigma_{ln\,\gamma}$ and σ_ω are statistically significant in the dynamic model (at the significance level of 5%), this is not the case in the static model. Also, the estimate of $\overline{\omega}$ in the static model is surprisingly high in comparison with the corresponding value in the dynamic model and it is not significant at the usual significance levels, which does not occur in the dynamic model. In addition, the estimates of the average efficiency across firms is relatively different in these two models: 1.31 for the dynamic model and 1.23 for the static model.

These differences between the estimates of the dynamic and static models raise the issue of model selection, that is, which model is more accurate to explain firm behaviour given our data. We use the Myopia test - a Likelihood Ratio Test (LR Test) - to know which model is more accurate. The usage of this classical test is based on

	$\delta = 0.975$			$\delta = 0$		
	Coefficient	St. Error	$P >\mid z - test \mid$	Coefficient	St. Error	$P >\mid z - test \mid$
$\overline{\beta^T}$	1.4106249	0.0374866	0.000	1.3333737	0.0336378	0.000
σ_{β^T}	0.2098812	0.0196476	0.000	0.1853932	0.0181456	0.000
$\overline{ln\,\gamma}$	-2.7797691	0.8217394	0.001	-6.5964089	6.6550349	0.322
$\sigma_{ln\,\gamma}$	0.7922932	0.2787332	0.005	0.9865202	0.7710622	0.201
$\overline{\omega}$	0.0370870	0.0277537	0.182	5.5886660	9.8630319	0.571
σ_{ω}	0.2064477	0.0217431	0.000	2.6976548	2.6086462	0.301
$\overline{\beta}$	1.4232952			1.3486076		
σ_{β}	0.1966898			0.1704699		
$\overline{\gamma}$	0.0849320			0.0022210		
σ_{γ}	0.0793717			0.0028499		
Obs.	1671			1671		
Log-L	-1090.547013			-1122.170305		
LR test	$\chi^2(1) = 63.246582$					
	$P > \chi^2(1) = 0.000$					

Table 2.5: Estimation Results - Small Firms

the fact that estimates obtained with the Bayesian method used in this paper converge to estimates obtained with the true posterior distribution of the parameters (see Imai *et al.* (2009)). In addition, the Bernstein-von Mises theorem establishes that estimates obtained with the true posterior distribution are asymptotically equivalent to estimates obtained using classical maximum likelihood procedures (e.g., see chapter 12 of Train (2009)). Thus, although the values of the likelihood function obtained with Bayesian and classic estimates may not be the same in finite samples, the two likelihoods have the same value asymptotically. Since all classical tests only have asymptotic validity, we are really no worse off when using our Bayesian estimates than when using classical maximum likelihood estimates.

The values of the log-likelihood function for both the dynamic and static models, as well as the LR test and the corresponding p-value, are shown in Table 2.3. The results clearly reject the hypothesis that the two models are identical, indicating that the dynamic model significantly improves our ability to fit the data.

Table 2.4 presents our estimation results using data only on big firms (532 observations). The results of the dynamic model show that average efficiency across big firms is around 1.41 and the maintenance cost parameter is approximately 679323 Euros, which is a very large cost. Interestingly, these estimates are bigger than the corresponding estimates for the entire dataset, that is, big firms are less efficient and have higher maintenance costs. Also, there seems to be more efficiency heterogeneity and maintenance costs heterogeneity across big firms. In addition, we found out that the estimates of $\overline{\gamma}$ and σ_γ are negligible.

The results in Table 2.4 for the dynamic and static models are quite different. This is true when we consider, for instance, the estimates of $\overline{\beta}$ and σ_β (respectively, 1.41 and 0.23 for the dynamic model and 1.24 and 0.15 for the static model), but the most significant difference is related to the estimates of $\overline{\omega}$ and σ_ω. In the dynamic model, the estimates of $\overline{\omega}$ and σ_ω are equal to 0.68 and 0.7 Millions of Euros, respectively, which are very high values, although probably not unrealistic (as the average replacement cost is still higher than these values - see Table 2.2). In the static model, the corresponding estimates are, respectively, around 625 and 24 Millions of Euros, which are unreasonably high values. Also, the static model predicts that $\overline{\omega}$ and σ_ω are not significant at the usual levels, which is not the case in the dynamic model. All these differences justify the use of the Myopia test, which is also presented in Table 2.4. The results indicate that the dynamic model is more capable of explaining big firms' behaviour.

Finally, Table 2.5 reports the results using data only on small firms (1671 observations). The results of the dynamic model show that average efficiency across small firms is around 1.42, which is a value very similar to the one found for big firms (1.41). However, this value is fairly different from the one found for the entire dataset (1.31), suggesting that independent estimation for different types of firms is important. The estimate of $\overline{\omega}$ is around 37087 Euros and it is not significant at the usual levels. However, the estimated σ_ω is statistically significant, suggesting that heterogeneity in maintenance costs across small firms is important. Interestingly, both estimates are smaller for small firms than for big firms or for the entire dataset, which seems to be reasonable. In addition, the estimated $\overline{\gamma}$ is around 84932 Euros, which is a relatively high value. In fact, this value is higher than the one for big firms (which is negligible) and for the entire dataset, possibly indicating that small firms have significant learning costs with a new machine.

The results in Table 2.5 for small firms also show some differences between the estimates of dynamic and the static model. In addition to the differences in, for instance, $\overline{\beta}$ and $\overline{\gamma}$, we have significant differences between the estimates of $\overline{\omega}$ and σ_ω in the two models. In particular, σ_ω is significant in the dynamic model, showing that heterogeneity matters, while it is not significant in the static model. The Myopia test reveals that dynamics is important as the dynamic model is better at explaining small firms' behaviour than the static model.

6 Counterfactual Experiments

We perform a counterfactual experiment to know what would be the firms' replacement choices if there were full efficiency in the decision process, that is, if β_i were equal to one for all firms and so actual costs were equal to minimum costs. To do this, we simulate the dynamic discrete choice model, using the estimates in Tables 2.3 to 2.5, with the exception of the estimates associated with β_i.

We simulate the model using 200 replications. The results are displayed in Table 2.6. We present results for the entire dataset, as well as for big and small firms separately. The "Actual Decisions" lines refer to the actual decisions made by firms, which are displayed in our dataset. The "Full Efficiency" lines refer to the counterfactual results, that is, the choices that firms would make in a full efficiency scenario (i.e., β_i is equal to one for all firms).

The results for the entire dataset show similar percentages of replacement for both actual and counterfactual scenarios (30.00% and 30.42%). This could lead us to conclude that firms' actual replacement decisions are not very different from the decisions in a full efficiency scenario. However, when we distinguish big and small firms, the percentages of actual and optimal decisions appear to be quite different. While big firms have decided to replace their machines in 31.02% of the decisions, they should have made this decision in 47.56% of the observations in order to be fully efficient. As for small firms, they have decided to do a replacement in 29.68% of the observations, although they should have done it only in 21.26% of the decisions.

Our results show that big firms should have replaced more machines while small firms should have decided not to do as many replacements as they did. These are perhaps not very surprising results if we take into account our estimates in Table 2.4 and 2.5. We saw that, apart from the net replacement cost, big firms bear negligible costs when they replace a machine, although they have significant maintenance costs. On the other hand, maintenance costs for small firms are not that significant, while their learning costs can be important. Thus, it is not very surprising that efficient big firms do many replacements and efficient small firms do not replace many machines.

We also analyze the impact of changes in replacement costs on the probability of replacement and compare actual changes in the probability of replacement with the

		All Firms	Big Firms	Small Firms
Actual	% Replacements	30.00%	31.02%	29.68%
Decisions	% No Replacements	70.00%	68.98%	70.32%
Full	% Replacements	30.42%	47.56%	21.26%
Efficiency	% No Replacements	69.58%	52.44%	78.74%

Table 2.6: Results for the Counterfactual Experiment

ones we would have in a full efficiency scenario. To do so, we simulate the dynamic model with respect to replacement costs while keeping the other state variables at their sample means and the random parameters at their estimated means presented in Tables 2.3 to 2.5.

In Figures 2.1 to 2.3 we graph how the probability of replacement changes as the replacement cost varies. The "Actual Estimates" curve presents results based on the actual estimates displayed in Tables 2.3 to 2.5 while the "Full Efficiency" curve is based on the same estimates, except for β_i, which is set equal to 1.

In Figure 2.1, we present the results for the entire dataset. Clearly, changes in replacement costs affect the probability of replacement, as when the replacement cost increases until 6 Millions of Euros, the probability of replacement decreases from around 60% to zero. Also, except for very low values of the replacement cost, the probability of replacement in the full efficiency hypothesis is always higher than the corresponding probability evaluated at the actual estimates of the parameters. This suggests that, for a significant range of the replacement cost, the efficient average firm is more likely to replace its machine than the actual average firm.

In Figure 2.2 we show the results using data only on big firms. Comparing to the previous case, the probability of replacement is higher for any value of the replacement cost. Also, it is not convex on the replacement cost for a fair range of it. Interestingly, the actual probability of replacement is much more sensitive to replacement costs than the corresponding probability in a full efficiency scenario. In fact, when compared with the efficient average big firm, the actual average big firm overestimates the probability of replacement at low replacement costs and underestimates it at high replacement costs.

Finally, Figure 2.3 reports the results for small firms. The results are fairly similar to the ones obtained for the entire dataset. In this case, the probability of replacement is always higher in the full efficiency hypothesis than the corresponding probability evaluated at the actual estimates.

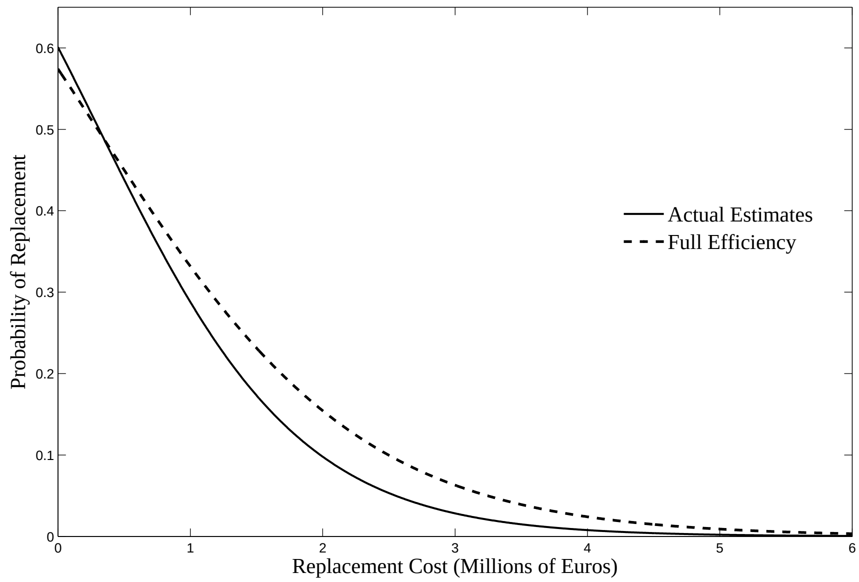

Figure 2.1: Probability of Replacement - All Firms

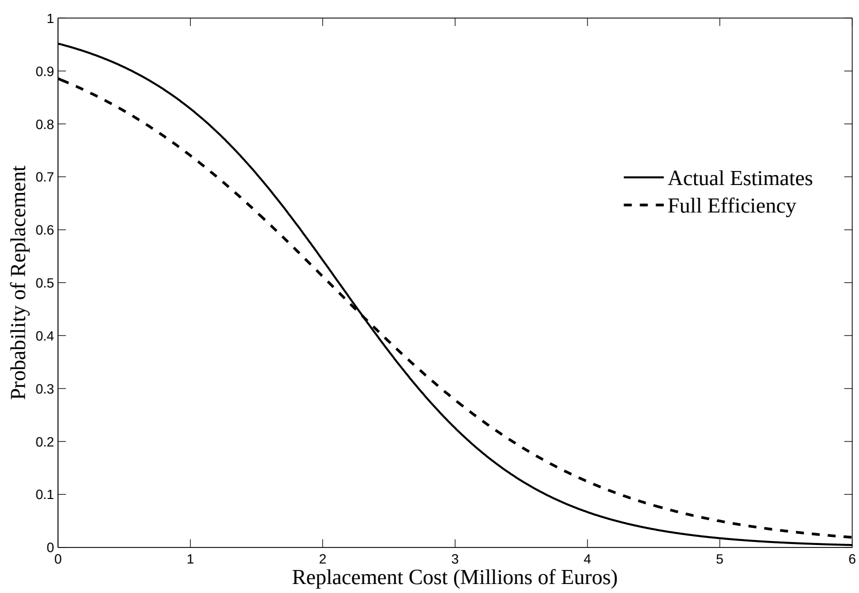

Figure 2.2: Probability of Replacement - Big Firms

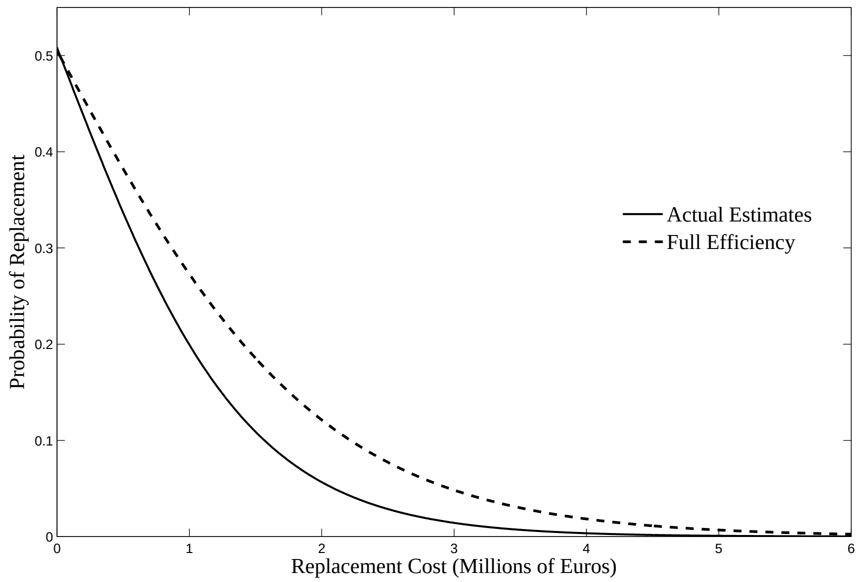

Figure 2.3: Probability of Replacement - Small Firms

7 Conclusion

In this paper, we use the dynamic discrete choice framework to analyze efficiency at the firm level. We develop a dynamic machine replacement model in which firms attempt to minimize intertemporal costs and introduce a measure of dynamic efficiency in this model. The model allows for the existence of firm heterogeneity as well as efficiency heterogeneity across firms by considering that the structural parameters are random parameters.

We use a dataset with yearly information for 290 Portuguese manufacturing firms from 2001 to 2008 and estimate the model for all the firms in our dataset, as well as for subsamples on big and small firms. We find out that the estimated efficiency is very similar for big and small firms as the average efficiency parameter is around 1.41 for big firms and 1.42 for small firms, which means that big and small firms' actual costs are 1.41 and 1.42 times their minimum costs, respectively. Our estimation results also reveal that different types of firms have different cost structures: on average, while big firms have significant maintenance costs and negligible learning costs, small firms bear important learning costs and statistically insignificant maintenance costs.

We perform counterfactual experiments to know what would be the firms' optimal choices and compare them with the actual choices. The counterfactuals results suggest that big firms should have done more replacements and small firms should have replaced less in order to be fully efficient. We also analyze the impact of changes in replacement costs on the probability of replacement and compare actual changes in the probability of replacement with the ones we would have in a full efficiency scenario. The results show that the actual probability of replacement for the average big firm is much more sensitive to replacement costs than the corresponding probability in a full efficiency scenario and that the probability of replacement for the average small firm is always higher in the full efficiency hypothesis than the corresponding probability evaluated at the actual estimates.

We believe that the use of the dynamic discrete choice framework to evaluate efficiency at the firm level is promising as it allows us to have information regarding firm efficiency that until now was only available for models with continuous decision variables. One of the major advantages of this approach is the possibility of performing counterfactual experiments, allowing for comparisons between actual decisions and optimal decisions. This analysis makes it possible for researchers to have more information regarding firm's decisions and payoffs, which hopefully will allow us to get a better understanding of firms' behaviour.

Appendix

A. Nonparametric Estimation of the Transition Probability of c_{it}

Following Sanchez-Mangas (2002), we generate a nonparametric estimate of $f_c(c_{i,t+1} \mid c_{it})$. We use $M = 26$ cells to discretize the state variable c_{it}. Let us denote the discretized values of the state variable by c_{it}^c and the values of the state variable in the m^{th} cell by c_{it}^m.

We estimate the transition probabilities for c as

$$\hat{Prob}(c_{t+1}^c = c^m \mid c_t^c = c^l) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} I\left\{c_{i,t+1}^c = c^m\right\} K_1\left(c_{it}, c^l\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T} K_1\left(c_{it}, c^l\right)},$$

for $m, l = 1, ..., 26$. K_1 is the univariate gaussian kernel defined as

$$K_1\left(c_{it}, c^l\right) = \frac{1}{(2\pi)^{1/2}} exp\left\{-\frac{1}{2}\left(\frac{c_{it} - c^l}{h_1}\right)^2\right\},$$

where h_1 is the bandwidth parameter, defined according to the Silverman's rule.

B. Computation of the Final Estimates of the Parameters

Here we show how to compute the final estimates of the means and standard deviations of β_i and γ_i from the estimated values in Tables 2.3 to 2.5 .

For $\overline{\beta^T}$ and σ_{β^T}, let us define the adjusted values by $\overline{\beta}$ and σ_β respectively. We can compute such values as (see Johnson *et al.* (1994))

$$\overline{\beta} = \overline{\beta^T} + \frac{\phi\left(\frac{1-\overline{\beta^T}}{\sigma_{\beta^T}}\right)}{1 - \Phi\left(\frac{1-\overline{\beta^T}}{\sigma_{\beta^T}}\right)} \times \sigma_{\beta^T}$$

and

$$\sigma_\beta = \sqrt{\sigma^2_{\beta^T} \times \left[1 - \frac{\phi\left(\frac{1-\overline{\beta^T}}{\sigma_{\beta^T}}\right)}{1 - \Phi\left(\frac{1-\overline{\beta^T}}{\sigma_{\beta^T}}\right)} \left(\frac{\phi\left(\frac{1-\overline{\beta^T}}{\sigma_{\beta^T}}\right)}{1 - \Phi\left(\frac{1-\overline{\beta^T}}{\sigma_{\beta^T}}\right)} - \frac{1-\overline{\beta^T}}{\sigma_{\beta^T}}\right)\right]},$$

where ϕ and Φ denote the standard normal probability density function and the standard normal cumulative distribution function respectively.

For $\overline{ln\,\gamma}$ and $\sigma_{ln\,\gamma}$, let us define the final values as $\overline{\gamma}$ and σ_γ respectively. The final mean and the final standard deviation are given by (see chapter 6 of Train (2009))

$$\overline{\gamma} = exp(\overline{ln\,\gamma} + \sigma^2_{ln\,\gamma}/2)$$

and

$$\sigma_\gamma = \sqrt{exp(2\overline{ln\,\gamma} + \sigma^2_{ln\,\gamma}) \times (exp(\sigma^2_{ln\,\gamma}) - 1)}.$$

References

[1] Aguirregabiria, V. and P. Mira (2002), "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models", Econometrica, 70, 1519-1544.

[2] Aguirregabiria, V. and P. Mira (2010), "Dynamic Discrete Choice Structural Models: A Survey", Journal of Econometrics, 156, 38-67.

[3] Ching, A., S. Imai, M. Ishihara and N. Jain (2010), "A Guide to Bayesian Estimation of Dynamic Discrete Choice Models with an Application to a Store-Level Reward Program". Available at SSRN: http://ssrn.com/abstract=1398444.

[4] Cooper, R., J. Haltiwanger and L. Power (1999), "Machine Replacement and the Business Cycle: Lumps and Bumps", American Economic Review, 89, 921-946.

[5] Das, S. (1992), "A Micro-Econometric Model of Capital Utilization and Retirement: The Case of the U.S. Cement Industry", Review of Economic Studies, 59,277-297

[6] Hotz, J. and R. Miller (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models", Review of Economic Studies, 60, 497-529.

[7] Imai, S., N. Jain and A. Ching (2009), "Bayesian Estimation of Dynamic Discrete Choice Models", Econometrica, 77, 1865-1899.

[8] Johnson, N., S. Kotz and N. Balahrishnan (1994), "Continuous Univariate Distributions", New York, 2nd edition.

[9] Kasahara, H. (2009), "Temporary Increases in Tariffs and Investment: The Chilean Case", Journal of Business and Economic Statistics, 27, 113-127.

[10] Kennet, M. (1994), "A Structural Model of Aircraft Engine Maintenance", Journal of Applied Econometrics, 9, 351-368

[11] Lasserre, P. and P. Ouellette (1999), "Dynamic Factor Demands and Technology Measurement under Arbitrary Expectations", Journal of Productivity Analysis, 11, 219-241.

[12] McFaddden, D. (1978), "Modelling the Choice of Residential Location", in: A. Karlquist *et al.*, (Eds.), Spatial Interaction Theory and Planning Models, North Holland, Amsterdam, 75–96.

[13] Magnac, T. and D. Thesmar (2002), "Identifying Dynamic Discrete Decision Processes", Econometrica, 70, 801-816.

[14] Miranda, M. and G. Schnitkey (1995), "An Empirical Model of Asset Replacement in Dairy Production", Journal of Applied Econometrics, 10, S41-S55

[15] Nemoto, J. and M. Goto (1999), "Dynamic Data Envelopment Analysis: Modeling Intertemporal Behavior of a Firm in the Presence of Productive Inefficiencies", Economic Letters, 64, 51-56.

[16] Nemoto, J. and M. Goto (2003), "Measurement of Dynamic Efficiency in Production: An Application of Data Envelopment Analysis to Japanese Electric Utilities", Journal of Productivity Analysis, 19, 191-210.

[17] Ouellette, P. and L. Yan (2008), "Investment and Dynamic DEA", Journal of Productivity Analysis, 29, 235-247.

[18] Rust, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher", Econometrica, 55, 999-1033.

[19] Rust, J. (1994), "Structural Estimation of Markov Decision Processes", in R. E. Engle and D. McFadden (Eds.), Handbook of Econometrics, 4, North-Holland, Amsterdam, 3081-3143.

[20] Rust, J. and G. Rothwell (1995), "Optimal Response to a Shift in Regulatory Regime: The Case of the US Nuclear Power Industry", Journal of Applied Econometrics, 10, S75-S118.

[21] Sanchez-Mangas, R. (2002), "Pseudo-Maximum Likelihood Estimation of a Dynamic Structural Investment Model", Working Paper 02-62, Statistics and Econometrics Series, Universidad Carlos III de Madrid.

[22] Serra, T., A. O. Lansink and S. Stefanou (2011), "Measurement of Dynamic Efficiency, a Directional Distance Function Parametric Approach", American Journal of Agricultural Economics, 93, 756-767.

[23] Silva, E. and S. Stefanou (2007), "Dynamic Efficiency Measurement: Theory and Application", American Journal of Agricultural Economics, 89, 398-419.

[24] Train, K. (2009), "Discrete Choice Methods with Simulation", Cambridge University Press, 2nd Edition.

Essay 3: Efficiency Measurement in Dynamic Discrete Choice Processes

Abstract

We use the discrete choice framework to analyze efficiency at the micro level, explicitly assuming that firms make decisions over discrete variables. We develop an efficiency measure for Single-Agent dynamic discrete choice models and for dynamic Empirical Games in which the payoff function may be the cost function, the revenue function or the profit function. The efficiency measure is illustrated with the estimation of a dynamic empirical game for the Portuguese banking industry between 2002 and 2009.

Keywords: Efficiency Measurement, Dynamic Discrete Choice Models, Strategic Interaction

JEL Classification: C15, C25, C61, D21

1 Introduction

There is a vast literature on firm efficiency measurement in a static framework (e.g., see Färe *et al.* (2004), Das and Kumbhakar (2010), Portela and Thanassoulis (2007) and Berger *et al.* (2007)). Recently, there has been some promising work on evaluating firm efficiency in a dynamic environment (e.g., see Silva and Stefanou (2007), Nemoto and Goto (1999, 2003), Lasserre and Ouellette (1999), Ouellette and Yan (2008) and Serra *et al.* (2011)).

Efficiency measurement at the firm level - in either static or dynamic environments - has been analyzed considering that firms make decisions over continuous variables. There are, however, some situations in which firms do make decisions over discrete variables, rather than over continuous variables. Examples include decisions on product ordering, patent renewal, machine replacement, price changes, as well as situations in which firms make decisions in the presence of indivisible inputs or outputs.

Since the seminal work of McFadden (1978) and its extension to Markov decision problems by Rust (1987), Hotz and Miller (1993) and Aguirregabiria and Mira (2002), among others, the discrete choice framework has been used in a wide variety of applications in which agents decide over discrete variables. Aguirregabiria and Mira (2010) presents an interesting survey on applications of discrete choice models in a dynamic environment.

We use the discrete choice framework to analyze firm efficiency considering that firms attempt to maximize intertemporal payoffs and make decisions over discrete variables. Inefficiency arises whenever firms deviate from their optimal choices, therefore getting an actual payoff smaller than the optimal payoff. Firstly, an efficiency measure is developed within a Single-Agent dynamic discrete choice model in which the payoff function may be the cost function, the revenue function or the profit function. Secondly, we extend our efficiency measure to dynamic Empirical Games, that is, discrete choice models in which there is strategic interaction among firms, meaning that each firm's payoff is affected by the other firms' decisions. We describe the Nested Pseudo Likelihood (NPL) Method proposed by Aguirregabiria and Mira (2002, 2007), which can be used to estimate Single-Agent Models and a particular class of Empirical Games. We also generalize the identification results of Aguirregabiria and Magesan (2010), which uses an extension of the NPL method to estimate more general Empirical Games.

An empirical illustration is presented involving the estimation of a dynamic empirical game and the efficiency measure for the Portuguese banking industry between 2002 and 2009. We consider that banks attempt to maximize the intertemporal revenue associated with the provision of each service. The results indicate that banks are, on average, revenue efficient, though the interaction among banks has different impacts for each service.

The paper is organized as follows. In Section 2, we develop the efficiency measure in a Single-Agent model followed by a discussion of the NPL method proposed by Aguirregabiria and Mira (2002). Section 3 discusses efficiency measurement in Empirical Games and develops a generalization of the identification results in Aguirregabiria and Magesan (2010) to a model with N firms and J choices. Additionally, an extension of the NPL method is discussed. The empirical illustration is developed in Section 4. Section 5 concludes the paper.

2 Single-Agent Models

2.1 A Single-Agent Model with Inefficiency

In a single-agent model, firms make decisions over discrete variables in order to maximize intertemporal payoffs. There is no strategic interaction among firms, thus a firm's payoff does not depend on the other firms' decisions.

Time is discrete and indexed by t, $t = 0,...,\infty$, while firms are indexed by i, $i = 1,...,$ N. The decision variables a_{it} belong to a discrete, finite set $A = \{0, 1, ..., J-1\}$. For firm i in a given time period t, we denote the observable state variables by $x_{it} \in \chi$, where χ has a finite support, and the unobservable (to the econometrician) state variables by ε_{it}. We assume that x_{it} has a discrete, finite support with transition probability function $f(x_{i,t+1} \mid a_{it}, x_{it})$, that is, beliefs about the next period values of x_{it} do not depend on ε_{it}. Also, we consider that ε_{it} are independent over time with distribution function $G(\varepsilon_{it})$.

Let us denote the current payoff function for firm i by

$$\Psi(a_{it}, x_{it}, \varepsilon_{it}; \theta') = \psi(a_{it}, x_{it}; \theta') + \varepsilon_{it}(a_{it}), \tag{3.1}$$

where θ' represents the vector of parameters. This function represents firm i's optimal payoff. A natural way to analyze efficiency in the model is to include an additive term in the current payoff function that ensures that optimal payoffs are greater than or equal to actual payoffs. This is an approach that is in the spirit of the Stochastic Frontier Analysis (e.g., see Kumbhakar and Lovell (2000)). However, in a discrete choice model "only differences in utility matter" (see Train (2009)) and such an inefficiency term would not be identifiable. Allowing for inefficiency, the current payoff function is redefined as

$$\Psi(a_{it}, x_{it}, \varepsilon_{it}; \theta') = \psi(a_{it}, x_{it}; \theta') + \rho + \varepsilon_{it}(a_{it}), \tag{3.2}$$

where ρ represents the payoff loss due to inefficiency. Since ρ is not identifiable, we define ρ as a function of several observable variables, that is, $\rho = (\Gamma(\beta) - 1)\, \psi(a_{it}, x_{it}; \theta')$. Substituting ρ in the previous equation yields

$$\Psi(a_{it}, x_{it}, \varepsilon_{it}; \theta) = \Gamma(\beta) \times \psi(a_{it}, x_{it}; \theta') + \varepsilon_{it}(a_{it}), \tag{3.3}$$

where $\Gamma(\beta)$ is a function of β, $0 < \beta \leq 1$, and $\theta = (\theta', \beta)$.

The specification of $\Gamma(\beta)$ depends on the payoff function. Let us focus on the component in (3.3) given by $\Gamma(\beta) \times \psi(a_{it}, x_{it}; \theta')$. Suppose that we are modeling the behaviour of firms that attempt to minimize costs and let $C(a_{it}, x_{it}; \theta')$ be the cost function. In the terminology of (3.3), this corresponds to define $\psi(a_{it}, x_{it}; \theta') = -C(a_{it}, x_{it}; \theta')$.[1] We define $\Gamma(\beta) = 1/\beta$ and the firm's actual costs can be written as $1/\beta \times C(a_{it}, x_{it}; \theta')$. If $\beta = 1$, the firm is efficient and optimal and actual costs coincide; if $0 < \beta < 1$, there is

[1]Since we consider likelihood-based estimation methods, it is more convenient to define $\psi(a_{it}, x_{it}; \theta) = -C(a_{it}, x_{it}; \theta)$ and maximize the likelihood function rather than using $C(a_{it}, x_{it}; \theta)$.

inefficiency and actual costs are greater than minimum costs.

Alternatively, we may consider that firms attempt to maximize revenue. We define $\psi(a_{it}, x_{it}; \theta') = R(a_{it}, x_{it}; \theta')$ as the (optimal) revenue function. If firms are inefficient, then the actual revenue is smaller than the maximum revenue. Thus, we define $\Gamma(\beta) = \beta$ and the actual revenue is given by $\beta \times R(a_{it}, x_{it}; \theta')$. If $\beta = 1$, firms are efficient and optimal and actual revenues coincide; if $0 < \beta < 1$, there is inefficiency and the actual revenue is smaller than the maximum revenue, that is, $\beta \times R(a_{it}, x_{it}; \theta') < R(a_{it}, x_{it}; \theta')$.

Finally, we may consider that firms attempt to maximize profits. We define the profit function as $\psi(a_{it}, x_{it}; \theta') = \pi(a_{it}, x_{it}; \theta')$. By definition, profits can be either positive or negative. To take into account the possible existence of negative profits, the function $\Gamma(\beta)$ assumes different values for positive and negative profits. If $\pi(a_{it}, x_{it}; \theta') \geq 0$, then we set $0 < \Gamma(\beta) = \beta \leq 1$. This ensures that actual profits are equal to maximum profits if the firm is efficient (i.e., $\beta = 1$). If there is inefficiency, then $0 < \beta < 1$ and actual profits are smaller than maximum profits: $\beta \times \pi(a_{it}, x_{it}; \theta') < \pi(a_{it}, x_{it}; \theta')$. If $\pi(a_{it}, x_{it}; \theta') < 0$, then we set $\Gamma(\beta) = 1/\beta \geq 1$. When there is no inefficiency ($\beta = 1$), actual profits are equal to maximum profits. If there is inefficiency, then $0 < \beta < 1$ and actual profits are smaller than maximum profits: $1/\beta \times \pi(a_{it}, x_{it}; \theta') < \pi(a_{it}, x_{it}; \theta')$, since $\pi(a_{it}, x_{it}; \theta') < 0$.

We have defined alternative payoffs and the corresponding functions $\Gamma(\beta)$. Given the state variables x_{it} and ε_{it}, the problem of the firms is to make decisions a_{it} in order to maximize the expected discounted payoffs over time

$$E\left(\sum_{t=0}^{\infty} \delta^t \, \Psi(a_{it}, x_{it}, \varepsilon_{it}; \theta)\right),$$

where $\Psi(a_{it}, x_{it}, \varepsilon_{it}; \theta)$ is defined in (3.3) and $\delta \in (0, 1)$ is the discount factor.[2] Let $V(x_{it}, \varepsilon_{it}; \theta)$ be the value function of the dynamic programming problem. Given the assumptions on the transition probabilities for the state variables, we can write the dynamic programming problem using the concept of integrated Bellman equation (see Aguirregabiria and Mira (2010))

$$\bar{V}(x_{it}; \theta) \equiv \int V(x_{it}, \varepsilon_{it}; \theta) \, dG(\varepsilon_{it}) = \int \underset{a \in A}{max} \left\{\overline{v}(a, x_{it}; \theta) + \varepsilon_{it}(a)\right\} dG(\varepsilon_{it}),$$

where $\overline{v}(a, x_{it}; \theta) \equiv \Gamma(\beta) \times \psi(a, x_{it}; \theta') + \delta E_x \left[\bar{V}(x_{it+1}; \theta) \mid a, x_{it}\right]$ is the choice-specific value function. The Conditional Choice Probability (CCP) can be defined as

[2] The static discrete choice model emerges here as a particular case of the dynamic model when $\delta = 0$.

$$P(a \mid x;\theta) = \int I\left\{\overline{v}(a, x_{it};\theta) + \varepsilon_{it}(a) > \overline{v}(a', x_{it};\theta) + \varepsilon_{it}(a'),\, \forall a'\right\} dG(\varepsilon_{it}). \tag{3.4}$$

2.2 Identification and Estimation

Rust (1994), Magnac and Thesmar (2002) and Aguirregabiria (2010) study identification in single-agent models and prove that the dynamic discrete choice model is nonparametrically nonidentified. However, under some conditions - namely assuming a parametric specification for the current payoff function, $\varepsilon_{it}(a_{it})$ enters additively in the model and the transition probabilities for x_{it} and ε_{it} are independent from each other - it is possible to identify the parameters of the model (Rust (1994), Magnac and Thesmar (2002) and Aguirregabiria (2010)). The discount factor is the only parameter that remains underidentified. Thus, in general, the discount factor is not estimated and is assumed to be known by the researcher.

Several methods have been proposed to estimate dynamic discrete choice models (e.g., see Aguirregabiria and Mira (2010)).[3] We briefly describe the Nested Pseudo Likelihood (NPL) Method proposed by Aguirregabiria and Mira (2002), which is a likelihood-based method. The NPL method has been widely used in empirical analysis, namely in applications on labor economics and industrial organization (e.g., see Aguirregabiria and Mira (2010) for references on applications of the NPL method in several fields of microeconomics).

We consider that the discount factor δ and the distribution function $G(\varepsilon_{it})$ are known to the researcher and that the transition probabilities for x_{it}, $f(x_{i,t+1} \mid a_{it}, x_{it})$, have already been estimated (e.g., using Kernel methods or simply a frequency estimator). Let us denote the vector of CCPs by $\boldsymbol{P}(\theta) = \{P(a \mid x;\theta) : (a, x) \in A \times \chi\}$ and the vector of value functions by $\bar{\boldsymbol{V}}(\theta) = \left\{\bar{V}(x;\theta) : x \in \chi\right\}$. Having data on $i = 1, ..., N$ firms during $t = 1, ..., T$ periods, the (conditional) likelihood function is defined as

$$L(\boldsymbol{P};\theta) = \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=0}^{J-1} P(a \mid x;\theta)^{I\{a=j\}}. \tag{3.5}$$

Note that the likelihood function in (3.5) depends on the CCP defined in (3.4), which in turn depends on the unknown value function. The choice probabilities defined

[3] Some of the most popular methods include Rust (1987)'s Nested Fixed Point Algorithm, Hotz and Miller (1993)'s CCP method, Aguirregabiria and Mira (2002)'s Nested Pseudo Likelihood Method and Imai *et al.* (2009)'s Bayesian Method.

in (3.4), $P(a \mid x; \theta)$, can be written in vector form as $\boldsymbol{P}(\theta) = \Upsilon(\bar{\boldsymbol{V}}(\theta))$, where $\Upsilon(.)$ maps a vector in the value function space into a vector of CCPs. Hotz and Miller (1993) proves that the mapping in (3.4) - which relates choice probabilities, $P(a \mid x; \theta)$, and differences in choice-specific value functions, $\tilde{v}(x_{it}; \theta) = \{\overline{v}(a, x_{it}; \theta) - \overline{v}(0, x_{it}; \theta) : a \in A\}$ - is invertible. This means that the value functions can be written in terms of conditional choice probabilities, which in vector form can be defined as $\bar{\boldsymbol{V}}(\theta) = \varphi(\boldsymbol{P}(\theta))$, where $\varphi(.)$ maps a vector of CCPs into a vector in the value function space. Therefore, we can write the vector of CCPs as

$$\boldsymbol{P}(\theta) = \Upsilon(\bar{\boldsymbol{V}}(\theta)) = \Upsilon(\varphi(\boldsymbol{P}(\theta))) \equiv \Lambda(\boldsymbol{P}; \theta), \tag{3.6}$$

showing that (3.4) can be expressed as a fixed point problem in the probability space.

Given the results of Hotz and Miller (1993), Aguirregabiria and Mira (2002) proposes an iterative procedure - known as the NPL Algorithm - to estimate the parameters in a dynamic (single-agent) discrete choice model. Given an initial consistent estimator for $\boldsymbol{P}$, say $\hat{\boldsymbol{P}}^{(0)}$, the NPL algorithm at iteration K allows us to obtain $\hat{\theta}^{(K)}$ and $\hat{\boldsymbol{P}}^{(K)}$ by iterating on

$$\hat{\theta}^{(K)} = arg\,\max_{\theta}\ L(\hat{\boldsymbol{P}}^{(K-1)}; \theta) \tag{3.7}$$

and

$$\hat{\boldsymbol{P}}^{(K)} = \Lambda(\hat{\boldsymbol{P}}^{(K-1)}; \hat{\theta}^{(K)}). \tag{3.8}$$

Note that, given (3.6), the likelihood function no longer depends on the value functions, as the CCPs can be written using (3.6). Aguirregabiria and Mira (2002) proves that the estimator $\hat{\theta}_{NPL}$ - obtained by iterating on (3.7) and (3.8) until convergence is achieved - is (root-n) consistent, asymptotically normal and asymptotically equivalent to the (partial) maximum likelihood estimator.

3 Empirical Games

3.1 An Empirical Game with Inefficiency

As in a single-agent model, in an empirical discrete game firms make decisions over discrete variables in order to maximize intertemporal payoffs. However, in an empirical game there is strategic interaction among firms, meaning that each firm's payoff is affected by the other firms' decisions.

Time is discrete and indexed by t, $t = 0,...,\infty$, while firms are indexed by i, $i = 1,..., N$. The decision of firm i at period t is represented by a_{it}, which belongs to a discrete, finite set $A = \{0, 1, ..., J-1\}$ and $a_t = (a_{1t}, a_{2t}, ..., a_{Nt})$. For a given time period t, we denote the observable state variables by $x_t \in \chi$, $x_t = (x_{1t}, x_{2t}, ..., x_{Nt})$ and the unobservable state variables (which can be interpreted as firms' private information) by $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})$. We assume that x_t has a discrete, finite support with transition probability function $f(x_{t+1} \mid a_t, x_t)$, that is, beliefs about the next period values of x_t do not depend on ε_t. Also, we consider that ε_{it} are independent over time and across firms with distribution $G(\varepsilon_{it})$.

Let us denote the current payoff function for firm i by

$$\Psi(a_{it}, a_{-it}, x_t, \varepsilon_{it}; \theta') = \psi(a_{it}, a_{-it}, x_t; \theta') + \varepsilon_{it}(a_{it}), \tag{3.9}$$

where a_{-it} represents the actions of all players (other than i) in period t and θ' represents the vector of parameters. This function represents firm i's optimal payoff. As in single-agent models, we analyze efficiency using the function $\Gamma(\beta)$, with $0 < \beta \leq 1$. Allowing for inefficiency, the current payoff function is redefined as

$$\Psi(a_{it}, a_{-it}, x_t, \varepsilon_{it}; \theta) = \Gamma(\beta) \times \psi(a_{it}, a_{-it}, x_t; \theta') + \varepsilon_{it}(a_{it}), \tag{3.10}$$

where $\theta = (\theta', \beta)$.

As before, the specification of $\Gamma(\beta)$ depends on the definition of the payoff function. If we consider that firms attempt to minimize costs, we define $\Gamma(\beta) = 1/\beta$ and $\psi(a_{it}, a_{-it}, x_t; \theta') = -C(a_{it}, a_{-it}, x_t; \theta')$. Alternatively, if firms attempt to maximize the revenue, then we define $\Gamma(\beta) = \beta$ and $\psi(a_{it}, a_{-it}, x_t; \theta') = R(a_{it}, a_{-it}, x_t; \theta')$. Finally, we may consider that firms attempt to maximize profits and, in this case, we define $\psi(a_{it}, a_{-it}, x_t; \theta') = \pi(a_{it}, a_{-it}, x_t; \theta')$. If $\pi(a_{it}, a_{-it}, x_t; \theta') \geq 0$, then we set $\Gamma(\beta) = \beta$; if $\pi(a_{it}, a_{-it}, x_t; \theta') < 0$, then $\Gamma(\beta) = 1/\beta$.

Firm i's payoff depends on the actions taken by the other firms. Since the firms' decisions in a given time period are taken simultaneously, firm i does not know the other firms' decisions. Let $\sigma_i(x_t, \varepsilon_{it}; \theta)$ be the strategy function of firm i. Given this strategy function, the CCP for firm i represents firm i's actual behaviour and can be defined by

$$P_i(a \mid x; \theta) = \int I\{\sigma_i(x_t, \varepsilon_{it}; \theta) = a\} dG(\varepsilon_{it}). \tag{3.11}$$

Note that firm i does not observe the actual behaviour of the other firms when making its decision in a given time period, that is, firm i does not know $\sigma_j(x_t, \varepsilon_{jt}; \theta)$ and $P_j(a \mid x; \theta)$, $\forall j \neq i$. Thus, we define firm i's beliefs about the behaviour of the other firms. Let $b_{ij}(x_t, \varepsilon_{jt}; \theta)$ represent firm i's beliefs about the strategy function of firm j. The CCP associated with beliefs $b_{ij}(x_t, \varepsilon_{jt}; \theta)$ is given by

$$B_{ij}(a \mid x; \theta) = \int I\{b_{ij}(x_t, \varepsilon_{jt}; \theta) = a\} dG(\varepsilon_{jt}). \tag{3.12}$$

Given (3.12), firm i's current expected payoff is defined by

$$E_\Psi(a_{it}, x_t, \varepsilon_{it}; \theta) = \sum_{a_{-it} \in A^{N-1}} \left(\prod_{j \neq i} B_{ij}(a_{-it}[j] \mid x; \theta) \right) \Psi(a_{it}, a_{-it}, x_t, \varepsilon_{it}; \theta), \tag{3.13}$$

where $a_{-it}[j]$ represents the jth firm's element in a_{-it} and $A^{N-1} = \prod_{j \neq i} A$.[4] Given our definition of $\Psi(a_{it}, a_{-it}, x_t, \varepsilon_{it}; \theta)$ in (3.10), we can write (3.13) as

$$E_\Psi(a_{it}, x_t, \varepsilon_{it}; \theta) = \sum_{a_{-it} \in A^{N-1}} \left(\prod_{j \neq i} B_{ij}(a_{-it}[j] \mid x; \theta) \right) (\Gamma(\beta) \times \psi(a_{it}, a_{-it}, x_t; \theta') + \varepsilon_{it}(a_{it}))$$

$$= \Gamma(\beta) \times \left[\sum_{a_{-it} \in A^{N-1}} \left(\prod_{j \neq i} B_{ij}(a_{-it}[j] \mid x; \theta) \right) \psi(a_{it}, a_{-it}, x_t; \theta') \right] + \varepsilon_{it}(a_{it})$$

$$= \Gamma(\beta) \times E_\psi(a_{it}, x_t; \theta') + \varepsilon_{it}(a_{it}), \tag{3.14}$$

where

$$E_\psi(a_{it}, x_t; \theta') = \sum_{a_{-it} \in A^{N-1}} \left(\prod_{j \neq i} B_{ij}(a_{-it}[j] \mid x; \theta) \right) \psi(a_{it}, a_{-it}, x_t; \theta'). \tag{3.15}$$

Given the state variables (x_t and ε_{it}) and beliefs $B_{ij}(.)$, $\forall j \neq i$, the problem of firm i is to make decisions a_{it} in order to maximize the expected discounted payoffs over time $E\left(\sum_{t=0}^{\infty} \delta^t E_\Psi(a_{it}, x_t, \varepsilon_{it}; \theta)\right)$, where $E_\Psi(a_{it}, x_t, \varepsilon_{it}; \theta)$ is defined in (3.14) and $\delta \in (0, 1)$ is the discount factor.[5] Let $V_i(x_t, \varepsilon_{it}; \theta)$ be the value function of the dynamic

[4] For instance, if there are only 2 firms, firm i and firm j, then a_{-it} only includes the actions for firm $j \neq i$, that is, $a_{-it} = a_{jt}$. In this case, (3.13) is equivalent to $E_\Psi(a_{it}, x_t, \varepsilon_{it}; \theta) = \sum_{a_{jt}=0}^{J-1} B_{ij}(a_{jt} \mid x; \theta)\Psi(a_{it}, a_{-it} = a_{jt}, x_t, \varepsilon_{it}; \theta)$.

[5] The static empirical game emerges here as a particular case of the dynamic game when $\delta = 0$.

programming problem. Given the assumptions on the transition probabilities for the state variables, we can write the dynamic programming problem using the concept of integrated Bellman equation (see Aguirregabiria and Mira (2010))

$$\bar{V}_i(x_t;\theta) \equiv \int V_i(x_t,\varepsilon_{it};\theta)\, dG(\varepsilon_{it}) = \int \max_{a\in A} \{\overline{v}_i(a,x_t;\theta) + \varepsilon_{it}(a)\}\, dG(\varepsilon_{it}), \tag{3.16}$$

where $\overline{v}_i(a,x_t;\theta) \equiv \Gamma(\beta) \times E_\psi(a,x_t;\theta') + \delta E_x\left[\bar{V}_i(x_{t+1};\theta) \mid a, x_t)\right]$ is the choice-specific value function and the expectation $E_x(.)$ is taken over the distribution

$$E_f(x_{t+1} \mid a_{it}, x_t) = \sum_{a_{-it}\in A^{N-1}} \left(\prod_{j\neq i} B_{ij}(a_{-it}[j] \mid x;\theta)\right) f(x_{t+1} \mid a_t = (a_{it}, a_{-it}), x_t).$$

The CCP for firm i - which reflects firm i actual behaviour as defined in (3.11) - can be defined as

$$P_i(a \mid x;\theta) = \int I\left\{\overline{v}_i(a,x_t;\theta) + \varepsilon_{it}(a) > \overline{v}_i(a',x_t;\theta) + \varepsilon_{it}(a'),\ \forall a'\right\} dG(\varepsilon_{it}). \tag{3.17}$$

3.2 Identification and Estimation

Identification and estimation of empirical games have been analyzed assuming that beliefs are in equilibrium, that is, $B_{ij}(a \mid x;\theta) = P_j(a \mid x;\theta)$, $\forall x$ and $\forall i \neq j$ (e.g., see Aguirregabiria and Mira (2007), Bajari *et al.* (2007), Pakes *et al.* (2007) and Pesendorfer and Schmidt-Dengler (2008)). To the best of our knowledge, Aguirregabiria and Magesan (2010) is the only study that investigates identification and estimation of empirical games without imposing that restriction.

Aguirregabiria and Magesan (2010) presents identification results for a model with 2 firms and 2 actions. We extend their results to the general model with N firms and J choices. For comparability purposes, we discuss identification using the setup of Aguirregabiria and Magesan (2010), that is, considering that $\Psi(a_{it} = 0, a_{-it}, x_t, \varepsilon_{it};\theta) = 0$. As discussed in Aguirregabiria and Magesan (2010), the identification results do not depend on this specific normalization of the payoff function. In fact, any other normalization of the payoff function could be used since in discrete choice models "only differences in utility matter".

Let us assume that $f(x_{t+1} \mid a_t, x_t)$ and δ are known and consider the following assumptions:

Assumption 1 *(Exclusion Restriction): Let* $x_t = (w_t, s_{1t}, s_{2t}, ..., s_{Nt})$*, where* w_t *represents a vector of state variables common to all firms and* s_{it} *represents firm i's specific state variable. Firm i's payoff function,* $\Psi(a_{it}, a_{-it}, x_t, \varepsilon_{it}; \theta)$*, does not depend on* s_{jt}*,* $\forall j \neq i$*, that is,*

$$\Psi(a_{it}, a_{-it}, x_t, \varepsilon_{it}; \theta) = \Psi(a_{it}, a_{-it}, w_t, s_{it}, \varepsilon_{it}; \theta).$$

Assumption 2 *(No Strategic Uncertainty at J 'extreme' points): For every firm j, there are J values for* s_{jt} *-* $s^1_{jt}, s^2_{jt}, ..., s^J_{jt}$ *- such that for every value of* $(w_t, s_{1t}, \cdots, s_{j-1t}, s_{j+1t}, ..., s_{Nt})$ *the beliefs are in equilibrium, that is,* $\forall z = 1, ..., J$ *and* $\forall i \neq j$ *we have*

$$B_{ij}(a \mid w_t, s_{1t}, \cdots, s_{j-1t}, s^z_{jt}, s_{j+1t}, ..., s_{Nt}; \theta) = P_j(a \mid w_t, s_{1t}, \cdots, s_{j-1t}, s^z_{jt}, s_{j+1t}, ..., s_{Nt}; \theta).$$

Assumption 3 *(No Difference in Information among Rivals): The beliefs of firm i about the strategy function of firm j are equal to those made by firm k,* $i \neq k \neq j$*, that is, for any firm j we have*

$$B_{1j}(a \mid x; \theta) = \cdots = B_{j-1j}(a \mid x; \theta) = B_{j+1j}(a \mid x; \theta) = \cdots = B_{Nj}(a \mid x; \theta) = B_j(a \mid x; \theta).$$

Assumption 1 is a standard assumption in empirical dynamic games (e.g., see Pesendorfer and Schmidt-Dengler (2008)) and states that, although a firm's payoff may depend on the other firms' actions, it does not depend on the other firms' specific state variable. Pesendorfer and Schmidt-Dengler (2008) shows that if beliefs are always in equilibrium, Assumption 1 is a sufficient condition to guarantee nonparametric identification of the model.

Assumption 2 is a generalization of Aguirregabiria and Magesan (2010)'s "No Strategic Uncertainty at 2 'extreme' points" assumption. According to this assumption, it is recognized that firms' beliefs may not be always in equilibrium - that is, there are situations in which $B_{ij}(.) \neq P_j(.)$ - but there are some 'extreme' values for each firm's state variable for which beliefs are in equilibrium, that is, there is no strategic uncertainty. Aguirregabiria and Magesan (2010) shows that Assumptions 1 and the "No Strategic Uncertainty at 2 'extreme' points" assumption are sufficient to nonparametrically identify the model with 2 firms and 2 actions.

Assumption 3 establishes that firm i's beliefs about the behaviour of firm j may not be in equilibrium, but those beliefs are not different from the beliefs made by any other rival of firm j. All the firms (other than firm j) have the same information regarding firm j and so have the same beliefs about its behaviour.

Given Assumptions 1-3, Proposition 1 establishes our identification result for the general model with N firms and J actions:

Proposition 1 *Suppose there is sufficient variability in the data and that $f(x_{t+1} \mid a_t, x_t)$ and δ are known. Then, under Assumptions 1-3, the model is identified.*

The proof of Proposition 1 is in appendix A.

In order to estimate the model, we use an extension of the NPL method. Aguirregabiria and Mira (2007) extends the NPL method from single-agent models to dynamic games; Aguirregabiria and Magesan (2010) analyzes a model with 2 firms and 2 actions and extends the NPL method to estimate a dynamic game in which beliefs may not be in equilibrium.

We consider an estimation procedure in the context of a parametric payoff function and follow Aguirregabiria and Magesan (2010) in the details of the estimation method. Let us denote the vector of CCPs by $\boldsymbol{P}(\theta) = \{P_i(a \mid x;\theta) : (a,x) \in A \times \chi, \forall i = 1,...,N\}$ and the vector of beliefs by $\boldsymbol{B}(\theta) = \{B_i(a \mid x;\theta) : (a,x) \in A \times \chi,, \forall i = 1,...,N\}$.

Having data on $i = 1,...,N$ firms during $t = 1,...,T$ periods, the (conditional) likelihood function is defined as

$$L(\boldsymbol{P},\boldsymbol{B};\theta) = \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=0}^{J-1} P_i(a \mid x;\theta)^{I\{a=j\}}. \tag{3.18}$$

Note that the likelihood function in (3.18) depends on the CCP and beliefs defined in (3.11) and (3.12), but does not depend on the value function, since we can use the Hotz and Miller (1993)'s invertibility property in a similar way as we did in single-agent models.

Given an initial consistent estimator for $\boldsymbol{P}$ and $\boldsymbol{B}$, say $\hat{\boldsymbol{P}}^{(0)}$ and $\hat{\boldsymbol{B}}^{(0)}$, the extension of the NPL algorithm allows us to obtain $\hat{\theta}^{(K)}$, $\hat{\boldsymbol{P}}^{(K)}$ and $\hat{\boldsymbol{B}}^{(K)}$ at iteration K by performing the following 4 steps:

(i) Update of the structural parameters:

$$\hat{\theta}^{(K)} = arg\,\underset{\theta}{max}\;\; L(\hat{\boldsymbol{P}}^{(K-1)}, \hat{\boldsymbol{B}}^{(K-1)}; \theta) \tag{3.19}$$

(ii) Update of CCP functions:

$$\hat{\boldsymbol{P}}^{(K)} = \Lambda(\hat{\boldsymbol{P}}^{(K-1)}; \hat{\theta}^{(K)}) \tag{3.20}$$

(iii) Update of value functions: use Hotz and Miller (1993)'s invertibility property to write the differences in choice-specific value functions, $\tilde{v}_i(x_{it};\theta) = \{\overline{v}_i(a, x_{it};\theta) - \overline{v}_i(0, x_{it};\theta) : a \in A\}$, as a function of the CCPs and then use (3.16) to update the value functions. For instance, if we have a model with 2 choices ($a_{it} = 0$ and $a_{it} = 1$) and ε_{it} follows an iid extreme value distribution, then the integrated value function for a given firm i will be equal to

$$\bar{V}_i^{(K)}(x_t;\theta) = -log(1 - P_i^{(K)}(a = 1 \mid x;\theta)) + \overline{v}_i^{(K)}(0, x_{it};\theta). \tag{3.21}$$

(iv) Update of beliefs: let us define

$$q_i(a \mid x;\theta) = G^{-1}(P_i(a \mid x;\theta)) - \delta E'_x \left[\bar{V}_i(x_{t+1};\theta) \mid a, x\right],$$

where the expectation $E'_x(.)$ is taken over the distribution $E_f(x_{t+1} \mid a, x_t) - E_f(x_{t+1} \mid a = 0, x_t)$. Once the CCPs and value functions are set, we know the values for $q_i(a \mid x;\theta)$, that is, at iteration K we have

$$q_i^{(K)}(a \mid x;\theta) = G^{-1}(P_i^{(K)}(a \mid x;\theta)) - \delta E'_x \left[\bar{V}_i^{(K)}(x_{t+1};\theta) \mid a, x\right]. \tag{3.22}$$

Given the definition of $q_i(a \mid x;\theta)$, we can write

$$q_i(a \mid x;\theta) = \Gamma(\beta) \times \left[E_\psi(a_{it} = a, x;\theta') - E_\psi(a_{it} = 0, x;\theta')\right],$$

that is,

$$q_i(a \mid x;\theta) = \sum_{a_{-it} \in A^{N-1}} \left(\prod_{j \neq i} B_j(a_{-it}[j] \mid x;\theta)\right) \left[\Gamma(\beta) \times (\psi(a_{it} = a, .) - \psi(a_{it} = 0, .))\right],$$

and defining $\psi'(a_{it} = a, a_{-it}, x;\theta') = \Gamma(\beta) \times (\psi(a_{it} = a, a_{-it}, x;\theta') - \psi(a_{it} = 0, a_{-it}, x;\theta'))$, we have

$$q_i(a \mid x;\theta) = \sum_{a_{-it}\in A^{N-1}} \left(\prod_{j\neq i} B_j(a_{-it}[j] \mid x;\theta)\right) \psi'(a_{it} = a, a_{-it}, x; \theta'). \tag{3.23}$$

Given the parametric specification for the current payoff function, the structural parameters and the value for the state variables, $\psi'(a_{it} = a, a_{-it}, x; \theta')$ in (3.23) is known.

Let us define $\mathbf{q}(\theta) = \{q_i(a \mid x;\theta) : (a,x) \in A_{-0} \times \chi, \forall i = 1,...,N\}$, where $q_i(a \mid x;\theta)$ is defined in (3.23), $A_{-0} = A \backslash \{0\}$, and denote the dimension of the space χ by $\mid \chi \mid$. Given (3.23), $\mathbf{q}(\theta)$ represents a system of $(J-1) \times N \times \mid \chi \mid$ equations on $(J-1) \times N \times \mid \chi \mid$ beliefs, say $\mathbf{q}(\theta) = \Xi(\mathbf{B}(\theta))$.

Under Assumptions 1-3, we update beliefs at iteration K as follows:

- for those values of beliefs such that Assumption 2 applies, we set for every firm j

$$B_j^{(K)}(a \mid ., s_{jt}^z;\theta) = P_j^{(K)}(a \mid ., s_{jt}^z;\theta). \tag{3.24}$$

- for all the other values of beliefs we make use of the relationship in (3.23) and get the values for beliefs from the values of $\mathbf{q}(\theta)$, that is,

$$\mathbf{B}^{(K)} = \Xi^{-1}(\mathbf{q}^{(K)}(\theta^{(K)})). \tag{3.25}$$

By iterating on *(i)* to *(iv)*, we obtain the estimator $\hat{\theta}^*$, which is a (root-n) consistent and asymptotically normal estimator (see Aguirregabiria and Magesan (2010)).

4 Empirical Illustration

4.1 The Model

We illustrate the efficiency measure with the estimation of a dynamic empirical game for the Portuguese banking industry. We consider that, in every period, banks provide several services and choose whether to charge a regular or discount price for each service in order to maximize the expected discounted revenue over time associated with each service. We have information that allows us to compute the revenue associated with each service, but we are not able to compute the cost of each service. The available information on costs consider a bank as a whole and there is no detailed information regarding the cost of each service separately.

For a given service and time period, each bank may charge either a regular price or a discount price. Banks prefer to charge regular prices rather than discount prices as regular prices allow them to get a greater revenue. However, if other banks charge discount prices for the same service, a bank charging a regular price perceives a loss in sales because some of its clients may decide to buy the service at other banks. Similarly, a bank that charges a discount price for a given service perceives a gain in sales when other banks are charging regular prices for the same service. Thus, for each bank, the revenue associated with a given service depends on the price decisions of the other banks, meaning that there is strategic interaction among banks.

For a given bank i, service m and time period t, the decision variable a_{imt} belongs to $A = \{0, 1\}$, where $a_{imt} = 1$ if a regular price is charged and $a_{imt} = 0$ if a discount price is charged. Let $R(a_{imt}) \equiv R(a_{imt}, a_{-imt}, D_{imt}, \varepsilon_{imt}; \theta)$ denote the optimal current revenue function. We define this function as

$$R(a_{imt}) = p_{mt} D_{imt}(1 - \alpha_m I\{a_{imt} = 1\} \sum_{j \neq i} I\{a_{jmt} = 0\}$$

$$+ \gamma_m I\{a_{imt} = 0\} \sum_{j \neq i} I\{a_{jmt} = 1\}) + \varepsilon_{imt},$$

where p_{mt} represents the expected price charged for service m in time period t conditional on the decision variable a_{imt}, D_{imt} denotes potential sales of bank i for service m at period t, $I\{.\}$ is the indicator function and ε_{imt} is the random component, which we consider to be independent over time with type 1 extreme value distribution.

Actual sales may be different from potential sales D_{imt} if banks charge different prices for the same service, since some clients are expected to move from banks that are charging a regular price to banks charging a discount price for the same service. Thus, we include in the specification of the optimal revenue function the service-specific parameters α_m and γ_m to take into account differences in prices across banks. The parameter α_m measures the percentage loss in sales faced by a bank charging a regular price for service m when one more rival is charging a discount price for that service. Similarly, the parameter γ_m measures the percentage gain in sales obtained by a bank charging a discount price for service m when one more rival is charging a regular price for that service.

Allowing for inefficiency, the current revenue function is redefined as:

$$R(a_{imt}) = \beta_m p_{mt} D_{imt}(1 - \alpha_m I\{a_{imt} = 1\} \sum_{j \neq i} I\{a_{jmt} = 0\}$$
$$+ \gamma_m I\{a_{imt} = 0\} \sum_{j \neq i} I\{a_{jmt} = 1\}) + \varepsilon_{imt} \tag{3.26}$$

and $0 < \beta_m \leq 1$ is the service-specific inefficiency measure.

As for the state variable D_{imt}, we define a parametric specification for its transition probability:[6]

$$D_{imt+1} = \mu_{1m} D_{imt} + \mu_{2m} a_{imt} + \mu_{3m} \sum_{j \neq i} I\{a_{jmt} = 1\} + \mu_{4m} \sum_{j \neq i} I\{a_{jmt} = 0\} + u_{it}, \tag{3.27}$$

where μ_{1m}, μ_{2m}, μ_{3m} and μ_{4m} are service-specific parameters that capture the effect of current potential sales, bank's choice and rivals' behaviour on D_{imt+1}. We define $u_{it} = u_i^* + u_{it}^{**}$, where u_i^* represents a bank fixed effect that is constant over time and u_{it}^{**} is an iid component. To take into account the fixed effect u_i^*, we estimate (3.27) using the Arellano-Bond estimator (Arellano and Bond (1991)).

4.2 The Specification of Outputs and the Data

Banking technology is characterized by multiple inputs and multiple outputs. The specification of inputs and outputs for banks is not consensual in the banking literature. There are two main approaches to measure outputs and inputs: the production approach and the intermediation approach (e.g., Berger and Humphrey (1992, 1997)). The former approach has been used by, for example, Kuussaari and Vesala (1995); the latter approach has been employed by Portela and Thanassoulis (2005), Feng and Serletis (2009), among others. There are some studies that employ both approaches for comparison purposes (e.g., Berger *et al.* (1997)); others employ both approaches to select the inputs and outputs (e.g., Das and Kumbhakar (2010)).

Under the production approach, banks are assumed to perform transactions and process financial documents or providing counseling services to customers. According to this approach, outputs are measured by the number and type of transactions processed over a given time period (e.g., number of deposit accounts, number of loans provided). Inputs under this approach are measured by physical variables (e.g., labor and capital)

[6] Given the relatively small number of observations in our dataset, discussed in the next Section, and the well-known curse of dimensionality underlying nonparametric methods, we consider a parametric specification.

or their corresponding costs. Interest costs are excluded because only the operational process is relevant. On the other hand, under the intermediation approach, banks are viewed as providers of financial intermediation services consisting in the collection of deposits and other liabilities and their application in interest-earning assets (e.g., loans, securities). Outputs, under this approach, are measured by loans and other major assets; inputs are measured by operational and interest costs. The role of deposits is controversial. Deposits can be viewed as an input since generate interest payments and investible funds. Yet, deposits can also be considered as an output since they are associated with services provided to depositors such as liquidity and safekeeping. Three variants of the intermediation approach – the user cost approach, the asset approach and the value-added approach – are used in the literature (e.g., Berger and Humphrey (1992)). All these approaches focus on the intermediation activity of banks and, in general, loans and other major assets of banks are measured as outputs. According to the user cost approach, a financial product is an output if the financial return on an asset is greater than the opportunity cost of the funds or if the financial costs of a liability are less than the opportunity cost; otherwise the financial product is considered an input. Under the asset approach, deposits and other liabilities as well as other resources (e.g., labor and capital) are treated as inputs and outputs include only the bank assets (e.g., Färe *et al.* (2004); Allen and Liu (2007); Berger *et al.* (2007)). In general, under the value-added approach, the major categories of deposits and loans are considered outputs because they contribute to a significant proportion of the value-added (e.g., Drake *et al.* (2006); Epure *et al.* (2011)).

Focusing on the intermediation role of banks, we adopt the asset approach to define the banking activity. Accordingly, we assume that banks provide three types of services: credit, financial services associated with their operational activity and securities portfolios. For notational simplicity, hereafter we denote these services as services 1, 2 and 3, respectively. Credit is defined by bank's loans and we use received commissions to compute the values for financial services. Securities portfolios include cash and balances at Central Banks, bonds, other fixed income securities, shares and other variable income securities.

Our database is a balanced panel data with 360 observations that includes yearly information for 15 Portuguese commercial banks from 2002 to 2009. The data is taken from several issues of the *Boletim Informativo*, a yearly publication of the Portuguese Banking Association (*Associação Portuguesa de Bancos*) that gathers information at the bank level and covers virtually 100% of the banking activity in Portugal.[7] We dis-

[7]In 2005, the information in *Boletim Informativo* is subjected to some changes in methodology, which may have an

	Service 1: Credit		Service 2: Financial Services		Service 3: Securities	
	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev
D	15.8827	19.8131	0.1546	0.2116	2.8213	4.4913
p	5.4450	2.3108	0.6011	0.3792	4.6996	5.5461
a	0.4250	0.4964	0.4167	0.4951	0.3000	0.4602

Table 3.1: Descriptive Statistics

card from our sample the information for banks whose main activity is not related with commercial banking (e.g., investment banks) and consider only commercial banks for which there is information between 2002 and 2009. The banks in our sample represent more than 92.5% of the total volume of transactions for commercial banks in Portugal in each year.

The information in the *Boletim Informativo* allows us to get the prices for each service as well as to compute each bank's potential sales. Prices are expressed in percentage and potential sales are measured in 1000 Millions of Euros. The price of credit is determined as the ratio of interest receivable and similar income over loans and the price of financial services is computed as the ratio of commissions over costumer accounts, loans and advances to costumers. The price of securities portfolios is calculated as the ratio of income from securities, profit minus losses from financial transactions and other operating income over securities portfolios. A regular price ($a_{imt} = 1$) is a price greater than or equal to the average price across banks for a given service and a discount price ($a_{imt} = 0$) is a price below that average. We define potential sales D_{imt} as $D_{imt} = T_{mt} \times ms_{imt-1}$, where T_{mt} is a measure of market size - defined as the total volume of transactions for service m in period t - and ms_{imt-1} is bank i's market share for service m in period $t-1$. Thus, we are assuming that if there are no differences in prices among banks, clients will buy the services to the same bank they did in the past as they have no incentives to move to another bank. In this case, each bank keeps its position in the market (i.e., its market share) and each bank's sales are defined by the bank's market share in the previous period and the total volume of transactions that the clients are willing to do in the current period.

Table 3.1 presents some descriptive statistics. Potential sales D_{imt} appear to be relatively volatile since the standard deviation for each service is high in comparison with the corresponding mean. Also, banks charge discount prices in more than 50% of the observations (regular prices are charged in 42.5%, 41.67% and 30% of the observations

effect on how some of the variables are computed. Since we use data before and after 2005, our results should be seen with some caution.

for services 1, 2 and 3, respectively).

4.3 Results

As for the transition probability of the state variable D_{imt}, we have taken advantage that $\sum_{j \neq i} I\{a_{jmt} = 1\} = 14 - \sum_{j \neq i} I\{a_{jmt} = 0\}$, which results from the fact that we have 15 banks in our dataset. Therefore, the transition probability of D_{imt} in (3.27) can be written as

$$D_{imt+1} = \mu_{1m} D_{imt} + \mu_{2m} a_{imt} + \mu_{3m} \left[14 - \sum_{j \neq i} I\{a_{jmt} = 0\} \right] + \mu_{4m} \sum_{j \neq i} I\{a_{jmt} = 0\} + u_{it}$$

and so the equation we use in estimation is

$$D_{imt+1} = \mu_{1m} D_{imt} + \mu_{2m} a_{imt} + \mu_{3m}^* + \mu_{4m}^* \sum_{j \neq i} I\{a_{jmt} = 0\} + u_{it},$$

where $\mu_{3m}^* = 14\mu_{3m}$ and $\mu_{4m}^* = \mu_{4m} - \mu_{3m}$.

Table 3.2 presents the results for the transition probability of potential sales for each service. All the coefficients have the expected signal and are statistically significant at the 5% significance level.

We use the values displayed in Table 3.2 to estimate the dynamic empirical game with the current revenue function given in (3.26). We do not estimate the discount factor, which we consider to be equal to 0.975. The coefficient estimates, bootstrapped standard errors and associated p-values are shown in Table 3.3. The results indicate that banks are, on average, revenue efficient as the estimate of β for each service is not significantly different from 1.

The estimates of γ and α are quite different for each service. As for service 1, when one more rival sets a regular price, there is an average increase of 0.18% in the revenue of a bank charging a discount price; on the other hand, when one more rival sets a discount price, there is an average decrease of 6.72% in the revenue of a bank charging a regular price. Thus, a bank charging a discount price for service 1 benefits from having some of its rivals setting regular prices; however, this gain on sales is less than the loss on sales faced by the bank when it charges a regular price and some of the rivals are charging discount prices. This means that, on average, banks have little benefit when they charge lower prices than its rivals but have heavy losses when they charge higher

	Service 1: Credit		Service 2: Financial Services		Service 3: Securities	
	Coefficient	St. Error	Coefficient	St. Error	Coefficient	St. Error
μ_1	1.0390	0.0006	0.9186	0.0041	0.7497	0.0022
μ_2	-0.4935	0.0909	-0.0011	0.0005	-0.0671	0.0282
μ_3^*	0.3395	0.0310	0.0335	0.0027	2.0338	0.0354
μ_4^*	-0.0526	0.0056	-0.0025	0.0001	-0.2429	0.0024

Table 3.2: Results for the Transition Probability of D_{imt}

	Service 1: Credit			Service 2: Financial Services			Service 3: Securities		
	Coefficient	St. Error	p-value	Coefficient	St. Error	p-value	Coefficient	St. Error	p-value
β	0.9992	0.0135	0.000	0.9759	0.1591	0.000	0.9963	0.0753	0.000
γ	0.0018	0.0002	0.000	0.0310	0.0118	0.009	1.0000	0.8712	0.251
α	0.0672	0.0047	0.000	0.0172	0.0045	0.000	0.1955	0.0012	0.000

Table 3.3: Estimation Results ($\delta = 0.975$)

prices than the rivals.

The results for service 2 are somewhat different from those obtained for service 1. In fact, the average increase in the revenue of a bank charging a discount price is now of 3.1%, while the average decrease in the revenue of a bank charging a regular price is equal to 1.72%. These results suggest that, on average, the benefit of a bank that charges lower prices than its rivals is not very different from the loss that the bank has when it charges higher prices than the rivals.

As for service 3, the estimates indicate that when one more rival sets a regular price, there is not a significant increase in the revenue of a bank charging a discount price. Conversely, when one more rival sets a discount price, there is an average decrease of 19.55% in the revenue of a bank charging a regular price. This is a big effect, which probably justifies the relatively low percentage of regular prices for service 3 in our data (30%).

We note that our results should be seen with some caution as we have a relatively small number of observations in our dataset. In fact, we only have 8 observations for each bank and 120 observations for each service. Given that we estimate the dynamic model with only 360 observations, this small number of observations may affect our results.

5 Conclusion

Firms make decisions over continuous as well as over discrete variables. While there have been many studies investigating firm efficiency within models in which firms make decisions over continuous variables, little attempt has been made to evaluate firm efficiency in situations in which firms make decisions over discrete variables.

In this paper, we explicitly consider that firms make decisions over discrete variables and develop an efficiency measure within the dynamic discrete choice framework. We firstly analyze how to measure firm efficiency within a Single-Agent model, in which firms's decisions do not have any effect on the other firms' payoff, meaning that there is no strategic interaction among firms. Then, we proceed to dynamic Empirical Games, in which strategic interaction among firms is explicitly considered, meaning that each firm's payoff is affected by the other firms' decisions. For both types of discrete choice models - with and without strategic interaction among firms - we discuss how to evaluate firm efficiency for different payoff functions (e.g., cost function, profit function).

We illustrate the efficiency measure with the estimation of a dynamic empirical game for the Portuguese banking industry between 2002 and 2009. We consider that banks attempt to maximize the expected discounted revenue over time associated with each service. The results indicate that banks are, on average, revenue efficient, though the interaction among banks has different impacts for each service.

Appendix A

Proof of Proposition 1

We follow Pesendorfer and Schmidt-Dengler (2008) and Aguirregabiria and Magesan (2010). Pesendorfer and Schmidt-Dengler (2008) shows that an empirical dynamic game is identified if the number of restrictions in the model is greater than or equal to the number of unknown parameters. This is because the equilibrium conditions of the model can be rephrased as a system of equations that are linear in the parameters. Then, the identification problem consists on comparing the number of equations of the system (the restrictions in the model) with the number of unknown parameters. Aguirregabiria and Magesan (2010) shows that, under Assumption 1 and the "No Strategic Uncertainty at 2 'extreme' points" assumption, the order condition for identification is satisfied for the model with 2 firms and 2 actions. We show that, under Assumptions 1-3, the order condition for identification also holds for the general model with N firms and J actions.

Suppose that $f(x_{t+1} \mid a_t, x_t)$ and δ are known or have been estimated elsewhere. Let us denote the state space of $x_t = (w_t, s_{1t}, s_{2t}, ..., s_{Nt})$, w_t and s_{it}, $\forall i$, as χ, ω and $\mathcal{S}$, respectively, so that we have $\chi = \omega \times \mathcal{S}^N$. Let $\mid \chi \mid$, $\mid \omega \mid$ and $\mid \mathcal{S} \mid$ denote the dimension or number of elements in the space χ, ω and $\mathcal{S}$, respectively.

Let us compare the number of restrictions in the model and the number of unknown parameters for a given firm i. The model restrictions are defined in $P_i(a \mid x; \theta)$. If we do not impose Assumptions 1-3, we have $(J-1) \mid \chi \mid$ restrictions, the number of unknown parameters in the payoff function is $J^{N-1}(J-1) \mid \chi \mid$ and the number of unknown parameters in beliefs is $(J-1)(N-1) \mid \chi \mid$. Thus, the model is not identified as the number of restrictions is smaller than the number of unknown parameters, i.e., $\mid \chi \mid < \left[J^{N-1} + (N-1)\right] \mid \chi \mid$.

If Assumption 1 holds, then the number of unknown parameters in the payoff function is now equal to $J^{N-1}(J-1) \mid \omega \mid\mid \mathcal{S} \mid$. However, the model is not identified as the number of restrictions is still smaller than the number of unknown parameters, that is,

$$\mid \chi \mid < (N-1) \mid \chi \mid + J^{N-1} \mid \omega \mid\mid \mathcal{S} \mid . \tag{3.28}$$

Note, however, that if beliefs are always in equilibrium, then the first term in the right-hand side of (3.28) vanishes and the model is identified as the number of restrictions would be greater than the number of unknown parameters: $\mid \chi \mid \geq J^{N-1} \mid \omega \mid\mid \mathcal{S} \mid$, which is satisfied if $\mid \mathcal{S} \mid \geq J$.

If we add Assumption 2, then the number of unknown parameters in beliefs is equal to $(N-1)\left[(J-1) \mid \chi \mid - J(J-1) \mid \omega \mid\mid \mathcal{S} \mid^{N-1}\right]$, but the general model remains not identified as the number of restrictions remains smaller than the number of unknown parameters, that is,

$$\mid \chi \mid < (N-1) \mid \chi \mid - (N-1) \mid \omega \mid\mid \mathcal{S} \mid \left[J \mid \mathcal{S} \mid^{N-2} - J^{N-1}\right]. \tag{3.29}$$

Note that if $N = J = 2$, as considered in Aguirregabiria and Magesan (2010), the model is identified as the relationship in (3.29) would be reduced to $\mid \chi \mid = \mid \chi \mid$.

Adding Assumption 3, the number of unknown parameters in beliefs becomes equal to $(J-1) \mid \chi \mid - J(J-1) \mid \omega \mid\mid \mathcal{S} \mid^{N-1}$. Given that the beliefs made about the behaviour of every firm j are equal among rivals, we only need information on one of firm j's rivals to identify the beliefs of the other firm j's rivals. This assumption allows to identify the beliefs about the behaviour of every firm. In this case, the order condition for identification is satisfied as the number of restrictions is no longer smaller than the number of unknown parameters, that is, we have

$$| \chi | \geq | \chi | - | \omega || \mathcal{S} | \left[J | \mathcal{S} |^{N-2} - J^{N-1} \right]. \tag{3.30}$$

The order condition for identification (3.30) is satisfied as long as $| \mathcal{S} | \geq J$. Thus, under Assumptions 1-3, and given that $| \mathcal{S} | \geq J$, the model with N firms and J actions is nonparametrically identified.

References

[1] Aguirregabiria, V. (2010), "Another Look at the Identification of Dynamic Discrete Decision Processes: An Application to Retirement Behavior", Journal of Business and Economic Statistics, 28, 201-218.

[2] Aguirregabiria, V. and A. Magesan (2010), "Estimation of Dynamic Discrete Games When Players' Beliefs are Not in Equilibrium", Working Paper, University of Toronto.

[3] Aguirregabiria, V. and P. Mira (2002), "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models", Econometrica, 70, 1519-1544.

[4] Aguirregabiria, V. and P. Mira (2007), "Sequential Estimation of Dynamic Discrete Games", Econometrica, 75, 1-53

[5] Aguirregabiria, V. and P. Mira (2010), "Dynamic Discrete Choice Structural Models: A Survey", Journal of Econometrics, 156, 38-67.

[6] Allen, J. and Y. Liu (2007), "Efficiency and Economies of Scale of Large Canadian Banks", Canadian Journal of Economics, 40, 225-244.

[7] Arellano, M. and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", Review of Economic Studies, 58, 277-297.

[8] Bajari, P., L. Benkard, and J. Levin (2007), "Estimating Dynamic Models of Imperfect Competition", Econometrica, 75, 1331-1370.

[9] Berger, A. and D. Humphrey (1992), "Measurement and Efficiency Issues in Commercial Banking", in: Z. Griliches, (Eds.), Output Measurement in the Service Sector, University of Chicago Press, 245-279.

[10] Berger, A. and D. Humphrey (1997), "Efficiency of Financial Institutions: International Survey and Directions for Future Research", European Journal of Operations Research, 98, 175-212.

[11] Berger, A., J. Leusner, and J. Mingo (1997), "The Efficiency of Bank Branches", Journal of Monetary Economics, 40, 141-162.

[12] Berger, A., Astrid A. Dick, Lawrence G. Goldberg, and Lawrence J. White (2007), "Competition from Large, Multimarket Firms and the Performance of Small, Single-Market Firms: Evidence from the Banking Industry", Journal of Money, Credit and Banking, 39, 331-368.

[13] Das, A. and Kumbhakar, S. C. (2010), "Productivity and Efficiency Dynamics in Indian Banking: an Input Distance Function Approach Incorporating Quality of Inputs and Outputs", Journal of Applied Econometrics, doi: 10.1002/jae.1183

[14] Drake, L., M. Hall, and R. Simper (2006), "The Impact of Macroeconomic and Regulatory Factors on Bank Efficiency: A Non-parametric Analysis of Hong-Kong's Banking System", Journal of Banking and Finance, 30, 1443-1466.

[15] Epure, M., K. Kerstens, and D. Prior (2011), "Bank Productivity and Performance Groups: A Decomposition Approach based upon the Luenberger Productivity Indicator", European Journal of Operational Research, 211, 630-641.

[16] Färe, R., S. Grosskopf, and W. Weber (2004), "The Effect of Risk-based Capital Requirements on Profit Efficiency in Banking", Applied Economics, 36, 1731-1743.

[17] Feng, G., and A. Serletis (2009), "Efficiency and Productivity of the US Banking Industry, 1998-2005: Evidence from the Fourier Cost Function Satisfying Global Regularity Conditions", Journal of Applied Econometrics, 24, 105-138.

[18] Hotz, J. and Miller, R. (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models", Review of Economic Studies, 60, 497-529.

[19] Imai, S., N. Jain, and A. Ching (2009), "Bayesian Estimation of Dynamic Discrete Choice Models", Econometrica, 77, 1865-1899.

[20] Kumbhakar, S. and K. Lovell (2000), "Stochastic Frontier Analysis", Cambridge University Press, New York.

[21] Kuussaari, H., and J. Vesala (1995), "The Efficiency of Finnish Banks in Producing Payment and Account Transactions", Working paper 6/95, Bank of Finland.

[22] Lasserre, P. and P. Ouellette (1999), “Dynamic Factor Demands and Technology Measurement under Arbitrary Expectations”, Journal of Productivity Analysis, 11, 219-241.

[23] Magnac, T. and D. Thesmar (2002), “Identifying Dynamic Discrete Decision Processes”, Econometrica, 70, 801-816.

[24] McFaddden, D. (1978), “Modelling the Choice of Residential Location”, in: A. Karlquist *et al.*, (Eds.), Spatial Interaction Theory and Planning Models, North Holland, Amsterdam, 75–96.

[25] Nemoto, J. and M. Goto (1999), “Dynamic Data Envelopment Analysis: Modeling Intertemporal Behavior of a Firm in the Presence of Productive Inefficiencies”, Economic Letters, 64, 51-56.

[26] Nemoto, J. and M. Goto (2003), “Measurement of Dynamic Efficiency in Production: An Application of Data Envelopment Analysis to Japanese Electric Utilities”, Journal of Productivity Analysis, 19, 191-210.

[27] Ouellette, P. and L. Yan (2008), “Investment and Dynamic DEA”, Journal of Productivity Analysis, 29, 235-247.

[28] Pakes, A., M. Ostrovsky and S. Berry (2007), “Simple Estimators for the Parameters of Discrete Dynamic Games (with entry/exit examples)”, Rand Journal of Economics, 38, 373-399.

[29] Pesendorfer, M. and P. Schmidt-Dengler (2008), “Asymptotic Least Squares Estimators for Dynamic Games”, Review of Economic Studies, 75, 901-928.

[30] Portela, M.C.A.S., and Thanassoulis, E. (2005), “Profitability of a Sample of Portuguese Bank Branches and its Decomposition into Technical and Allocative Components”, European Journal of Operational Research, 162/3, 850-866.

[31] Portela, M.C.A.S. and Thanassoulis, E. (2007), ”Developing a Decomposable Measure of Profit Efficiency Using DEA”, Journal of the Operational Research Society 58, 481-490.

[32] Rust, J. (1987), “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher”, Econometrica, 55, 999-1033.

[33] Rust, J. (1994), “Structural Estimation of Markov Decision Processes”, in: R. E. Engle and D. McFadden, (Eds.), Handbook of Econometrics, Vol. 4, North-Holland, Amsterdam, 3081-3143.

[34] Serra, T., A. O. Lansink and S. Stefanou (2011), "Measurement of Dynamic Efficiency, a Directional Distance Function Parametric Approach", American Journal of Agricultural Economics, 93, 756-767.

[35] Silva, E. and S. Stefanou (2007), "Dynamic Efficiency Measurement: Theory and Application", American Journal of Agricultural Economics, 89, 398-419.

[36] Train, K. (2009), "Discrete Choice Methods with Simulation", Cambridge University Press, 2nd Edition.