Warsaw School of Economics
Institute of Econometrics
Department of Applied Econometrics

Department of Applied Econometrics Working Papers
Warsaw School of Economics
Al. Niepodleglosci 164
02-554 Warszawa, Poland

Working Paper No. 5-11

A microeconometric analysis of album sales success in the Polish music market

Mateusz Myśliwski
Warsaw School of Economics

This paper is available at the Warsaw School of Economics
Department of Applied Econometrics website at: http://www.sgh.waw.pl/instytuty/zes/wp/

A microeconometric analysis of album sales success in the Polish music market*

Mateusz Myśliwski†

Abstract

The aim of this paper is twofold. Firstly, we attempt to investigate the challenges for the constantly changing music industry on the example of Poland, positing a conclusion that both artists and lables could profit from a precisely determined set of factors influencing the ultimate sales success. On the other hand, the article intends to fill the gap between record industry analyses an econometric literature, as in the course of research, we found that the use of quantitative methods is rarely encountered in such analyses.

The study uses a self-compiled dataset, containing information on 619 albums, which appeared on the Official Sales Chart (OLiS) between 2008 and 2009. We propose three models for different quantitative variables and summarize the obtained results, stating that the use of microeconometric methods in this area of research seems promising.

Keywords: music industry, discrete choice models, partial proportional odds model

JEL Classification: C25, C51, C52, L82

*I would like to thank Agata Głogowska-Mikołajczyk for language correction and prof. Marek Gruszczyński for useful remarks while writing the Polish version of the paper.

†Warsaw School of Economics (SGH). E-mail: mateusz.mysliwski@gmail.com. Dataset available upon request. All the usual caveats apply.

1 Introduction

Realizing that or not, directly or indirectly, everyone takes part in shaping the music industry. A wide array of possible activities includes some obvious ones, like attending concerts or purchasing records from favourite artists. However, even expressing fondness for a particular tune might influence the supply and demand mechanism.

Among the existing market studies, descriptive ones tend to prevail, whereas the use of mathematical and statistical tools is not as widespread as in other sectors. One could distinguish two major research directions – modelling general changes in sales volume or using microdata on individual songs or records to explain why some of them outperform others in the market.

Our analysis aims to fit into the second identified group, though it implicitly incorporates some of the findings from papers with *macro* approach (see e.g. those of Liebowitz [2003], Oberholzer and Strumpf [2007], Ordanini [2006]). Given the fact that the collected data originates in a ranking list, we built on a methodology developed by Bradlow and Fader [2001] and Lee, Boatwright, and Kamakura [2003], whose paper was devoted to the problem of forecasting the prelaunch sales of recorded music. The first authors use the data on songs that appeared on *Billboard's* Hot 100 chart and prove that their unique trajectories could be described by a set of five variables. Lee et al. [2003] base on *Billboard's* Top 200 albums chart. Subsequently, they estimate a rather sophisticated hierarchical Bayesian model, which enables the prediction of approximate sales figures.

To make the set of covariates in our models encompass as many characteristics as possible, we adopted some of the ideas presented by Harchaoui and Hamdad [2000], who construct a hedonic price model for classical music records, taking into consideration, for instance, the fact whether the album was issued by a major label.

Polish copyright collection societies and labels keep the exact numbers of album sales undisclosed. Hence, by assuming direct unobservability, instead of putting sales volume in the role of a dependent variable, quantitative variables representing the propensity to achieve commercial success are explained.

The main purpose of the presented paper is to find a set of determinants of album sales success. As music records are perceived to be almost infinitely heterogeneous, empirical generalization was used. We thus try to develop models, which could assist labels and their marketing departments by providing forecasts of albums' success (which in our case is expressed as receiving a sales certification award or reaching number one in weekly sales charts).

The remaining part of this paper is divided in four following sections: sec-

tion 2 is devoted to the sole purpose of outlining the situation of the Polish music market. In the third section we provide more detail on the dataset, describe the variables as well as explain briefly the theoretical framework of the models used. The penultimate section contains the presentation of empirical results. The conclusions and ideas for further research are to be found in the last part of the paper.

2 Current state of Polish music market

The phonographic industry, as a part of *creative economy* (Howkins [2001]) is currently undergoing rapid changes, mainly as a result of the fact that Internet has become increasingly accessible to the wider public. Consequently, more and more listeners tend to migrate away from the CD format, whereas the sales of digital music are grabbing a bigger market share. According to the IFPI[1] data, the sales of music in the form of digital files increased tenfold between 2004 and 2008, reaching over 20% of the entire market. This sharp surge partially makes up for the 24% fall in physical sales, though the overall worldwide trend remains down. Such a direction of development entails the problem of illegal file sharing, which exposes the music industry to substantial losses and is considered to be the most severe obstacle for further growth (Hui and Png [2003]).

2.1 Key figures

By comparison with the majority of other countries, the Polish recording industry seems to have behaved differently in the recent years. While the global sales value declined from $23 billion in 2004 to $18.4 billion in 2004, the Polish market saw a leap of almost 20%. The domestic phonographic industry is, however, very small in relation to the worldwide market and thus, even most unusual occurences in Poland do not affect global trends.
Measured in sales dollars, Poland is the world's 23rd market. In Europe itself, Polish sales amount to a mere 1.5% of the total value.
Another striking characteristic of the industry is the fact, that it remains dominated by physical media, whose sales make up 92% of labels' income. The absence of worldwide leaders in digital music sales (iTunes and Amazon) in the market might be perceived as the factor responsible for relatively lower sales of computer music files.

[1] International Federation of the Phonographic Industry.

2.2 Collective rights management

The composers' and performers' copyrights are protected by organizations acting on their behalf, which, apart from that, play a supervisory role in the industry. In Poland there are four copyright collecting agencies: ZAiKS (*Society of Stage Authors and Composers*), SAWP (*Association of Performing Artists*), STOART (*Union of Performers*) and ZPAV (*Polish Society of the Phonographic Industry*). The last organization mentioned is the Polish National Group of the IFPI, which holds responsibility, among others, for conducting airplay surveys, compiling sales charts as well as bestowing *gold, platinum* and *diamond record* awards upon best-selling albums and singles.

2.3 Market structure and segmentation

The Polish music market is an oligopoly, dominated by four key players, which are referred to as *major labels* or simply *majors*. They are subsidiaries of multinational recorded music companies – EMI, Sony, Universal and Warner. Their influence on the market is unarguable, though in the recent years their market share measured with concentration ratio CR_4 has gradually fallen from 76.42% in 2005 to just 65% in 2009. This remains in accordance with observations for the world market (Alexander [2002]). Independent record labels appear to keep winning a market share from majors, though their efforts are strenuous, since the "Big Four" labels possess incomparably greater financial resources. Thus they are by far less susceptible to operational risk while allocating funds for promotion of a particular artist. Furthermore, smaller producers of phonograms are in a way dependent on majors' distribution channels. Alexander [1994] estimated the cost of vertical integration to acquire the distribution network in the United States at $100 million. Obviously, in Polish conditions, this number would be several times smaller. Limited access to distribution channels, though moderately alleviated by the expansion of online shops, persists to be a vital barrier for potential entrants. Possibly, this is one of the reasons why only six independent Polish labels have a market share greater than one percent.
The market, according to IFPI's recommendations, can be divided into four following segments: domestic repertoire, international repertoire, compilations and soundtracks, and classical music. The segment of domestic repertoire is relatively the most competitive one, as the Herfindahl-Hirschmann Index amounted to 1602, which is below the threshold of 1800 set by the European Commission, which, when exceeded, indicates high concentration. For the second and third segment, the HHI was equal to – respectively – 2117 and 2351.1. It is worth mentioning, that as far as compliations and

soundtracks are concerned, this segment's share amounted to 27% in 2009 and was the highest amongst IFPI member countries.

2.4 Record pricing

Price is believed to be one of the key factors determining the demand for a certain product. We will hence examine, whether a music album can achieve commercial success just because it is substantially cheaper than other records. Mixon and Ressler [2000] conducted a study of price elasticity of demand for records, showing that the demand is inelastic as far as older albums are concerned. The demand for newly issued releases is, in turn, more elastic, since in some cases, people who have the desire to listen to new music tend to choose cheaper alternatives.
Even though the labels virtually have a monopolistic right to reproduce and publish the recordings, first degree price discrimination does not occur in the market. Instead, each record is assigned to a certain category on a price list called PPD (*Published Prices to Dealers*). Usually they are divided into domestic and foreign repertoire and within these further into: full, mid and budget prices. Which price range a record will belong to is determined by a handful of factors, that include the label's pricing strategy, estimated demand and sometimes, artist's decisions. In addition to that, end-user prices include a double-digit retailers' margin.

2.5 Sales certification

As mentioned in section 2.2, the Polish Society of the Phonographic Industry is responsible for certifying that a music recording has sold a certain number of copies. With a tradition dating back to the 1950s, the awards have the same names of *gold, platinum* and *diamond record* in almost all countries with IFPI National Groups.
In Poland, the first *gold records* were awarded in the 1960s, when sales of a long-playing record albums exceeded 160 thousand copies. In the 1990s, ZPAV was established and the hitherto monopolized market was freed from the shackles of central planning. A new set of thresholds was fixed at that time (100 and 50 thousand copies sold of, respectively, domestic and international repertoire). With the demand for recorded music gradually dropping, the criteria were liberalized twice – in 2003 and 2006. Current requirements are shown in table 1.
Two issues seem to require further explanation. Firstly, there is a possibility of being awarded with *multi-platinum* certification if the sales exceed the subsequent multiples of the platinum threshold. Secondly, the numbers

Table 1: Certification award levels in Poland.

	Gold	Platinum	Diamond
Domestic repertoire	15 000	30 000	150 000
International repertoire	10 000	20 000	100 000
Soundtracks	10 000	20 000	100 000
Jazz/classical	5 000	10 000	50 000

Source: Polish Society of the Phonographic Industry.

are counted for each disc the release comprises, and not for the release as a whole. The fact that jazz or classical music generally does not belong to the mainstream and the labels are not that willing to promote these music genres, underlies the slight diversity of the thresholds.
One ought to be aware that sales certification does not only serve the artist, for whom such an award might be an honour, but also yields valuable information about the success of an album, since the absolute sales figures are not made public. Chung and Cox [1994] use the number of *gold record* awards received as a measure of superstardom in their model, proving that it reflects the scale of success in a very similar way sales profit does.

3 Description of the data and theoretical framework

The following section gives details on the dataset and the data collection process, as well as provides brief information on the models used.

3.1 Data collection

The models presented in the next section were estimated with the data collected from OLiS (*Official Sales Chart*)[2]. The chart is compiled on the basis of reports received weekly from country's major retailers and intends to present a list of 50 countrywide best-selling albums each week. In our analysis, we take into consideration the period between January 2008 and December 2009 (102 weeks). A total of 663 different albums appeared on OLiS in the aforementioned time interval. However, we included only 619 observations in the estimation sample, for which a complete set of characteristics was possible to be obtained (i.e. albums that entered and left the charts in the analysed period)

[2]The charts can be accessed through ZPAV's official website: http://www.zpav.pl/index.php?lang=2

3.2 Variables

To present a wide range of approaches to the analysis of commercial success of an album, we estimate three models, each for a different dependent variable. Two of them are examples of binomial regression and thus the following discrete variables were modelled:

$$\text{NUMB1} = \begin{cases} 1 & \text{if the album reached the peak of OLiS charts} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{GOLD} = \begin{cases} 1 & \text{if the album got a } \textit{gold record} \text{ award} \\ 0 & \text{otherwise} \end{cases}$$

Following the binary response regressions, we propose an extension of the second model, where the dependent variable takes ordinal values, so that not only *gold records*, but also the two remaining certification awards are taken into consideration while predicting the probability of being awarded. The dependent variable used in the third model is thus as follows:

$$\text{AWARD} = \begin{cases} 3 & \textit{diamond record} \\ 2 & \textit{platinum record} \\ 1 & \textit{gold record} \\ 0 & \text{no award whatsoever} \end{cases}$$

We have refrained from the idea of adding extra tiers to the variable, corresponding to *multi-platinum* awards, as the newly created categories would be too infrequently represented.

The covariates set can be divided into two groups – the first one is a battery of five variables capturing chart trajectories for each album (Bradlow and Fader [2001]). It comprises the following measures: DEBUT – album's debut position (its rank at the time of the first appearance on the chart), WKS2PEAK – the number of weeks needed to reach the highest rank, PEAK – the top ranking the recording attains, TOTWKS – the total number of weeks spent on the list, EXIT – album's exit position (its rank during the last week of presence on the chart). To circumvent endogeneity and multicollinearity issues, we did not include all of these variables in the models and some of them were not used at all. However, we intend to present histograms for these five measures (see figure 1).

It can be observed that the histograms are somewhat similar in shape to those obtained by Bradlow and Fader [2001], even though they analysed songs, not albums. The majority of albums debuted on the charts below 15th place. Approximately 80% of them attained their peaks after one or two weeks.

There are some outliers – albums that needed more than 20 weeks to reach their top spot. This occured to records that returned in glory after leaving the charts and being absent for several weeks. Most albums exit the charts from bottom positions, which can be attributed to the fact that the loss of popularity is gradual. As far as the total number of weeks spent on the charts is concerned, over half of the albums did not *survive* for more than five weeks.
The second group of covariates comprises individual characteristics of the records. The rationale behind taking some of them into account has been already laid out in section 2. In this place we would only like to add that apart from variables defining the segment of the market it belongs to, album's retail price and whether the album was released under a major label, we analyse the influence of seasonal demand. Strobl and Tucker [2000] point out that its effect is likely to be significant and every year we observe that albums released on a special occasion (e.g. Christmas, various anniversaries, etc.) enjoy great popularity. Ultimately, we also consider the involvement of interactions (Cox [1984]). Table 2 summarizes this section.

Table 2: Explanatory variables.

Name	Possible values	Description
Variables regarding album's performance on the chart		
DEBUT	{1;50}	debut position
WKS2PEAK	{1;∞)	number of weeks needed to reach highest rank
PEAK	{1;50}	highest position attained
TOTWKS	{1;∞)	total number of weeks on the chart
EXIT	{1;50}	exit position
Other characteristics		
PRICE	continuous	average retail price
MAJOR	1 / 0	if the record was issued by a major label / otherwise
PL	1 / 0	if the album was recorded by a Polish artist / otherwise
VA	1 / 0	if the album is a compilation / otherwise
OST	1 / 0	if the album is a soundtrack or film score / otherwise
CLASSJAZZ	1 / 0	for jazz or classical music albums / otherwise
SEASON	1 / 0	if the album is a seasonal release / otherwise

3.3 Variable selection

As mentioned above, we were forced to refrain from including all the variables from table 2 in all the models because of endogeneity and multicollinearity issues. The first problem was avoided by analysing, to what extent adding

respective variables could be justified, in terms of sensible interpretation. For instance, using the PEAK covariate in the model for NUMB1 would not be justified. To detect possible multicollinearity and evaluate the strength of interdependencies between the variables, we analyse measures of association. Instead of computing a simple correlation matrix, which, in the case of ordinal and binary variables would be incorrect, we present an *interdependency matrix* (Gruszczyński [2002]). The idea of its construction is illustrated in table 3.

Table 3: The construction of an interdependency matrix.

Variable type	Continuous	Ordinal	Binary
Continuous	Pearson's rho	Spearman's rank correlation coeff.	*t*-test for difference in mean
Ordinal	Spearman's rank correlation coeff.		
Binary	*t*-test for difference in mean	Spearman's rank correlation coeff.	Yule's coeff. of association

Obviously, all measures but the *t*-test[3] show both the direction and the degree of the relationship. See table 12 for the computed interdependency matrix. Basically, the selection of covariates for our models could be summarized in three steps. Firstly, we check which variables yield good interpretation and inquire into measures of association to judge whether multicollinearity is an issue. Eventually, we estimate the models for the first time and either remove statistically insignificant covariates and reestimate the models, or leave some of the insignificant variables to prove our hypotheses right or wrong.

3.4 Theory underlying the models applied

The fact that each observation in our dataset corresponds to a micro-level entity and dependent variables are qualitative, leads to a conclusion that microeconometric methods are appropriate for this kind of analysis.
The existing literature on this subject provides extensive information on discrete choice models (see e.g. Cameron and Trivedi [2005], Maddala [1994], Long [1997]), so we will not elaborate on that. It seems worth mentioning that we estimated both logit and probit models in both cases and selected the better fitted one basing on information criteria. Treating the dependent variable as the observable result of a latent inclination to, respectively, get a *gold record* award or reach the peak of the charts, the probability of success given the set of covariates can be expressed as (1) assuming logistic distribution of the error terms and (2) if we assume normal distribution as in probit

[3]To denote statistically significant differences we use "+" (if the mean of the continuous variable is higher when the binary variable equals one) and "–" symbols. Otherwise, we put "0" in the matrix.

model:

$$p_i = \frac{e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}}} \tag{1}$$

$$p_i = \int_{-\infty}^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \tag{2}$$

For interpretation, we computed the effects of a discrete change for binary and ordinal variables and marginal effects for continuous covariates. To assess the quality of the model a battery of measures similar to the usual R^2 from the linear regression case is presented, though none of them yields an unequivocal interpetation in terms of percent of the explained variance. Long and Freese [2005] exhibit these goodness-of-fit measures in detail.

The third model we present is more complicated, since the dependent variable can take more than just two values, which can be put on an ordinal scale. Hence a multinomial model for the ordered outcomes seems suitable. In the very basic ordered logit model, introduced initially by Snell [1964] and developed further to include more than one covariate by McKelvey and Zavoina [1975], the coefficients are restricted not to vary across categories, so that the number of the parameters estimated is relatively small[4]. However, the model does not fall short and odds ratios are interpretable only under the rarely fulfilled *proportional odds assumption*. The cumulative probability that variable y is lower or equal to m conditioned by the covariates vector can be conveyed by:

$$\Pr(y_i \leq m|\mathbf{x_i}) = \frac{e^{\tau_m - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}}}{1 + e^{\tau_m - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}}} \quad \text{for } m \in \{1, ..., J-1\}. \tag{3}$$

It can be easily observed, that the parameters vector $\boldsymbol{\beta}$ stays the same regardless of which category of the variable we consider. Theoretically, (3) could be split into $J-1$ binary logistic regressions with different constant terms, so that their regression curves are parallel. If the assumption is violated[5], a different specification should be used. One of the solutions of this problem is estimation of a *generalized ordered logit model*, which allows category-varying parameters (Clogg and Shihadeh [1994]). Such model is often overparametrized, since we need to compute $k \times (J-1)$ coefficients.

To find a golden mean between an overparametrized and uninterpretable model, we suggest the estimation of a *partial proportional odds model*, introduced by Peterson and Harrel [1990]. A parsimonious specification is

[4] To be exact, k parameters for covariates and J-1 threshold values are estimated.

[5] A formal way of testing the proportional odds assumption has been proposed by Brant [1990].

achieved by imposing equality restrictions on as many parameters as possible, as long as the proportional odds assumption holds. Let $\mathbf{T_i}$ denote the part of the covariates vector, which includes variables causing the violation of the assumption and $\boldsymbol{\gamma}_m$ the corresponding vector of parameter estimates. Then, similiarily to (3), cumulative probability can be expressed as:

$$\Pr(y_i \leq m|\mathbf{x}_i) = \frac{e^{-\alpha_m - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{T}_i^{\mathrm{T}}\boldsymbol{\gamma}_m}}{1 + e^{-\alpha_m - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{T}_i^{\mathrm{T}}\boldsymbol{\gamma}_m}} \tag{4}$$

To estimate the parameters of the above equation we make use of Williams' [2005] algorithm `gologit2`. This model has also been recently described by Greene and Hensher [2011].

4 Results of the analysis

In the course of our research, many models were estimated. We arbitrarily chose to present three of them, which we consider to be the best ones, as far as goodness of fit and predictive power are concerned.

Logit model with GOLD as dependent variable

As one can observe in table 4, amongst the involved variables, only the price of the album does not influence the fact of it being awarded with *gold record*. Though the R^2 measures should be treated rather as benchmarks while com-

Table 4: Estimation results – logit model with GOLD as dependent variable.

Variable	Coefficient	Std. err.	*z*-statistic	*p-value*
Constant	-3.795612	0.7173013	-5.29	0.000 ***
TOTWKS	0.1863532	0.0252277	7.39	0.000 ***
DEBUT	-0.0343961	0.009941	-3.46	0.001 ***
MAJOR	1.920393	0.6935491	2.77	0.006 ***
PRICE	0.0040271	0.0084563	0.48	0.634
CLASSJAZZ	1.779997	0.4071556	4.37	0.000 ***
VA*SEASON	0.9477107	0.4089877	2.32	0.020 **

Number of observations:	619
Log likelihood:	-239.32
Pseudo-R^2:	0.328
McKelvey and Zavoina R^2:	0.553
Count-R^2:	0.801
Adj. Count-R^2:	0.369

Significance levels: *** 1% ** 5% * 10%

paring between different specifications, we can easily notice that their values significantly differ from zero. McKelvey and Zavoina R^2, which is said to resemble the usual coefficient of determination to the greatest extent, equals

0.553. Hence, we can infer that the model fits the data well. Since the coefficients of logistic regressions do not yield direct interpretation, odds ratios were computed (see table 5).
The obtained results are intuitive and easy to interpret. For instance, releasing the album under a major label increases the odds of it being awarded with *gold record* by 582.4%. The estimate corresponding to the interaction variable can be interpreted in a following manner: if the record is a seasonal compilation, the odds that it will sell well enough to receive the award are 2.5 times higher than in case of albums which are not seasonal compilations.

Table 5: Odds ratios.

Variable	Odds ratio	% change
TOTWKS	1.2048	20.5%
DEBUT	0.9662	-3.4%
MAJOR	6.8236	582.4%
PRICE	1.0040	0.4%
CLASSJAZZ	5.9298	493.0%
VA*SEASON	2.5798	158.0%

Eventually, we present several out-of-sample forecasts (see table 6), assuming different characteristics of the records. As for the in-sample predictive power, the count-R^2 shown in table 4 indicates that 80% of the forecasts are accurate.

Table 6: Out-of-sample forecasts.

Popular music album issued by a major					
TOTWKS	DEBUT	MAJOR	PRICE	CLASSJAZZ	VA*SEASON
average	average	1	average	0	0
$P(GOLD = 1\|\mathbf{x})$	0.2054	95% confidence interval:			[0.1618; 0.2489]
$P(GOLD = 0\|\mathbf{x})$	0.7946				[0.7511; 0.8382]
Popular music album issued by an independent label					
TOTWKS	DEBUT	MAJOR	PRICE	CLASSJAZZ	VA*SEASON
average	average	0	average	0	0
$P(GOLD = 1\|\mathbf{x})$	0.0365	95% confidence interval:			[-0.0108; 0.0838]
$P(GOLD = 0\|\mathbf{x})$	0.9635				[0.9162; 1.0108]
Classical/jazz album issued by a major					
TOTWKS	DEBUT	MAJOR	PRICE	CLASSJAZZ	VA*SEASON
average	average	1	average	1	0
$P(GOLD = 1\|\mathbf{x})$	0.6051	95% confidence interval:			[0.4280; 0.7822]
$P(GOLD = 0\|\mathbf{x})$	0.3949				[0.2178; 0.5720]
Seasonal compilation issued by a major, which debuted on 1st place					
TOTWKS	DEBUT	MAJOR	PRICE	CLASSJAZZ	VA*SEASON
average	1	1	average	0	1
$P(GOLD = 1\|\mathbf{x})$	0.6303	95% confidence interval:			[0.4154; 0.8452]
$P(GOLD = 0\|\mathbf{x})$	0.3697				[0.1548; 0.5846]

Observe, for example, that it is unlikely for a popular music album released under an independent label to get the *gold record* award (the probability of such an event is lower than 4%). The seemingly paradoxical fact that classical/jazz albums tend to be awarded relatively more frequently can be

attributed to two circumstances – firstly, the sales thresholds are lower for them (see table 1) and secondly, only albums that were successful enough to enter the charts were included in the sample. This hardly ever occurs to jazz or classical recordings, so those that even appeared on the charts were successfully enough to receive the award as well.

Probit model with NUMB1 as dependent variable

The second model is supposed to explain the album's inclination to reach number one position on the charts. By comparing information criteria for logit and probit models, we came to a conclusion that the latter is better. All

Table 7: Estimation results – probit model with NUMB1 as dependent variable.

Variable	Coefficient	Std. err.	*z*-statistic	*p-value*
Constant	-2.857557	0.3867971	-7.39	0.000 ***
TOTWKS	0.0666732	0.0123736	5.39	0.000 ***
PRICE	0.0116655	0.0065275	1.79	0.074 *
VA	-0.5315397	0.2684878	-1.98	0.048 **
PL*MAJOR	0.3502568	0.2144441	1.63	0.100 *

Number of observations:	619
Log likelihood:	-93.68
Pseudo-R^2:	0.234
McKelvey and Zavoina R^2:	0.288
Count-R^2:	0.825
Adj. Count-R^2:	0.091

Significance levels: *** 1% ** 5% * 10%

the included variables are statistically significant, though the direction of the relationship between the price and dependent variable seems counterintuitive and thus should not be interpreted. The quality of the model is satisfactory, though there is a high difference between count-R^2 and the adjusted count-R^2, which indicates that the predictive power is not particularly impressive due to unbalancedness of the sample.
Table 8 contains the estimates of marginal effects and discrete changes. Note that every additional week spent on the charts increases *ceteris paribus* the probability of reaching the peak by 0.0039. Furthermore, compliations are less likely to become number ones than albums including songs performed by just one artist.

Table 8: Marginal effects in the second model.

Variable	Min → Max	0 → 1	+/- $\frac{1}{2}$	Marginal effect
TOTWKS	0,9698	0,0015	0,0039	0,0039
PRICE	0,2061	0,0002	0,0007	0,0007
VA	-0,0254	-0,0254	-0,0318	-0,0307
PL*MAJOR	0,0243	0,0243	0,0205	0,0202

Several sample forecasts have been produced, for instance, the probability that a Polish album issued by a major label which stays on the charts for more than 10 weeks reaches number one is equal to 18%. On the contrary – for a compilation containing international repertoire, the probability is smaller than one percent.

Partial proportional odds model for AWARD

Finally, we present an extension of the first model that will also take into account the likelihood of receiving *platinum* and *diamond record* awards. Initially an ordered logit model was estimated and we ran a Brant test for parallel regressions. Null hypothesis, saying that the proportional odds assumption actually holds, needed to be rejected and thus we proceeded to estimating the PPOM. The results are shown in table 9.

Table 9: Estimation results – partial proportional odds model for AWARD.

Variable	Coefficient	Std. err.	*z*-statistic	*p-value*
PARAMETERS CONSTANT ACROSS CATEGORIES				
TOTWKS	0.1559843	0.0189896	8.21	0.000 ***
MAJOR	2.030631	0.5044222	4.03	0.000 ***
VA*SEASON	1.089625	0.3706633	2.94	0.003 ***
CLASSJAZZ	2.216956	0.4059533	5.46	0.000 ***
PARAMETERS VARYING ACROSS CATEGORIES				
Constant 0 (α_0)	-3.63058	0.5815562	-6.24	0.000 ***
Constant 1 (α_1)	-4.853446	0.6370371	-7.62	0.000 ***
Constant 2 (α_2)	-10.09953	1.083306	-9.32	0.000 ***
PEAK 0 (γ_{10})	-0.0472494	0.0118095	-4.00	0.000 ***
PEAK 1 (γ_{11})	-0.0970843	0.0243439	-3.99	0.000 ***
PEAK 2 (γ_{12})	-0.0138921	0.0377318	-0.37	0.713
Number of observations:	619			
Log likelihood:	-313.96			
Pseudo-R^2:	0.299			
Cragg & Uhler's R^2:	0.514			
Count-R^2:	0.787			
Adj. Count-R^2:	0.159			
Significance levels:	*** 1%	** 5%	* 10%	

PEAK was responsible for the fact that the parallel regressions assumption did not hold and hence, two additional parameters needed to be estimated. While interpreting the results, one needs to compare the odds across categories. We present the computed marginal effects in table 10.
As the number of various estimates is large, only a few interesting numbers will be given more attention. Notice, that due to the fact that there were only several *diamond records* in the sample, all the parameters for AWARD=3 are insignificant. For albums released by major labels, the probability of being awarded with *platinum record* is by 0.0238 higher than in the case when the same recording was issued under a smaller label. The chances of receiving

the *gold record* certification are definitely higher – by almost 0.16. Every additional week spent on the chart leads to an increase in probability of getting a *gold record* by 1.8 precentage points. For *platinum records* it is approximately 0.3 pp. We have already observed in the first model that seasonal compilations are by far more prone to get the *golden record* award. Now we are able to conclude, that they might have difficulties to reach the level of *platinum record*, as the estimate is on the verge of statistical significance.

Table 10: Marginal effects in the partial proportional odds model for AWARD.

Variable	Marginal effect	Std. err.	*z*-statistic	*p-value*
AWARD=0				
PEAK	0.0062413	0.001441	4.33	0.000 ***
TOTWKS	-0.0206045	0.0032936	-6.26	0.000 ***
MAJOR	-0.1845772	0.0281564	-6.56	0.000 ***
VA*SEASON	-0.1913487	0.0797698	-2.40	0.016 **
CLASSJAZZ	-0.4585632	0.0935471	5.46	0.000 ***
AWARD=1				
PEAK	-0.0044454	0.0013159	-3.38	0.001 ***
TOTWKS	0.017719	0.0029884	5.93	0.000 ***
MAJOR	0.159988	0.0260602	6.14	0.000 ***
VA*SEASON	0.158656	0.0640447	2.48	0.013 **
CLASSJAZZ	0.3416474	0.0633728	5.39	0.000 ***
AWARD=2				
PEAK	-0.0017879	0.0003924	-4.56	0.000 ***
TOTWKS	0.0027951	0.0010611	2.63	0.008 ***
MAJOR	0.0238233	0.0088015	2.71	0.007 ***
VA*SEASON	0.0316378	0.01919	1.65	0.099 *
CLASSJAZZ	0.1127974	0.0520206	2.17	0.030 **
AWARD=3				
PEAK	-0.000008	0.0000206	-0.39	0.696
TOTWKS	0.0000904	0.000063	1.43	0.151
MAJOR	0.0007659	0.0005585	1.37	0.170
VA*SEASON	0.0010548	0.0009237	1.14	0.253
CLASSJAZZ	0.0041184	0.0032273	1.28	0.202

Eventually, we will examine the predictive power of the last model. In-sample forecasts are illustrated in figure 2. For *diamond records* there were only three cases, where the predicted probability exceeded 0.2. Interestingly enough, we could be almost certain that some albums get the *platinum record* award, whereas for *gold records*, the situation seems slightly vague.
Out-of-sample forecasts similar to those in table 6 shall be presented now (see table 11).

One can easily observe, that for the first two hypothetical albums, the probability that they will not be successful enough to receive an award prevails. We have also computed the probabilities for some albums that actually did exist but were not in the estimation sample. For instance, for an internationally best-selling album by Nelly Furtado – 'Loose', we produced a forecast saying that it will with 98% certainty be awarded with a *diamond record*. This actually happened, which proves that the forecasts produced by the model can be of some use.

Table 11: Out-of-sample forecasts in the partial proportional odds model.

Popular music album issued by a major, which peaked at number 1				
PEAK	TOTWKS	MAJOR	VA*SEASON	CLASSJAZZ
1	average	1	0	0
No award	$P(AWARD = 0\|\mathbf{x})$	0,6306	95% confidence interval:	[0,5307; 0,7305]
Gold record	$P(AWARD = 1\|\mathbf{x})$	0,2284		[0,1490; 0,3079]
Platinum record	$P(AWARD = 2\|\mathbf{x})$	0,1400		[0,0754; 0,2046]
Diamond record	$P(AWARD = 3\|\mathbf{x})$	0,0009		[-0,0006; 0,0025]
Classical/jazz album issued by an independent label				
PEAK	TOTWKS	MAJOR	VA*SEASON	CLASSJAZZ
average	average	0	0	1
No award	$P(AWARD = 0\|\mathbf{x})$	0,7846	95% confidence interval:	[0,5790; 0,9902]
Gold record	$P(AWARD = 1\|\mathbf{x})$	0,1878		[0,0128; 0,3628]
Platinum record	$P(AWARD = 2\|\mathbf{x})$	0,0268		[-0,0088; 0,0623]
Diamond record	$P(AWARD = 3\|\mathbf{x})$	0,0009		[-0,0007; 0,0024]
Seasonal compilation issued by a major, which spent at least 10 weeks on the charts				
PEAK	TOTWKS	MAJOR	VA*SEASON	CLASSJAZZ
average	average above 10	1	1	0
No award	$P(AWARD = 0\|\mathbf{x})$	0,2151	95% confidence interval:	[0,0744; 0,3558]
Gold record	$P(AWARD = 1\|\mathbf{x})$	0,5108		[0,3717; 0,6499]
Platinum record	$P(AWARD = 2\|\mathbf{x})$	0,2628		[0,0615; 0,4641]
Diamond record	$P(AWARD = 3\|\mathbf{x})$	0,0113		[-0,0054; 0,0279]

5 Conclusions

The conducted study served two purposes – firstly, we intended to find the determinants of album sales success in the Polish music market. Secondly, we tried to verify whether microeconometric models can be used in such analyses. The empirical evidence we provided, corroborates that both goals were attained to a large extent.

We found that the price is an irrelevant factor and does not influence album's success, which confirms Liebowitz's [2003] findings. In accordance with our initial conjecture, major labels guarantee higher success probability. Moreover, compilations are extraordinarily popular in Poland and even if there are literally loads of them released every month, the chances that at least a fraction of them will turn out to be successful, are still higher than for most albums containing songs by just one artist.

The quality of the models presented is fairly good. One needs to be aware, that in case of microeconometric models, the goodness-of-fit assessment is not an obvious issue. The fact that the measures we presented were significantly greater than zero, and in some cases they exceeded 0.5, should serve as sufficient evidence that discrete response and multinomial models can be applied to such analyses.

However, we are fully aware of the fact, that our study is not entirely bereft of weaknesses. The sample encompassed only albums which appeared on the charts in a two-year period. Such an interval might seem too short to trace all the characteristics of the market. Furthermore, the covariates set we consider, can be questionned. On the one hand, it enables forecasting even for debutants, though on the other hand, the fact that we did not include some important characteristics like, for instance, the number of previously received

awards or label's marketing expenses, can be seen as a serious drawback of the analysis. Another question arises while considering, whether it makes any sense to analyse a market which, colloquially speaking, is doomed to extinction in a couple of years. The answer is not straightforward, though, even if physical media eventually cease to exist, forecasting the sales of digital files will remain an issue.

There is still a lot to be examined in the under-researched Polish and worldwide music industry. A similar study should be conducted for sales expressed in absolute values, as well as for other creative industries (e.g. book or film markets). An atheoretical analysis of albums' trajectories on the Polish charts is also worth considering. We hope that the conclusions derived from this paper will be useful for other researchers and labels' marketing managers.

References

Alexander P.J. [1994], *Entry barriers, release behavior, and multi-product firms in the music recording industry*, Review of Industrial Organization, **9**, pp. 85–98.

_____ [2002], *Market structure of the domestic music recording industry, 1890–1988*, Historical methods , **35**, pp. 129–132.

Bradlow E.T., Fader P.S. [2001], *A Bayesian lifetime model for the Hot 100 Billboard songs*, Journal of the American Statistical Association, **96**, pp. 368–381.

Brant R. [1990], *Assessing proportionality in the proportional odds model for ordinal logistic regression*, Biometrics, **46**(4), pp. 1171–1178.

Cameron A.C., Trivedi P.K. [2005], *Microeconometrics: methods and applications*, Cambridge University Press, New York.

Clogg, C., Shihadeh E. [1994], *Statistical models for ordinal variablepp. Advanced quantitative techniques in the social sciences series volume 4*, Sage Publications, Thousand Oaks.

Cox D.R. [1984], *Interaction*, International Statistical Review, **52**, pp. 1–31.

Chung K.H., Cox R.A.K. [1994], *A stochastic model of superstardom: an application of the Yule distribution*, The Review of Economics and Statistics, **76**, pp. 771–775.

Dixon R.D. [1982], *LP chart careers: Indices and predictors of ascent and descent in popularity*, Popular Music and Society, **8**, pp. 19–44.

Greene W.H., Hensher D.A. [2010]: *Modeling ordered choices. A primer*, Cambridge University Press, New York.

Gruszczyński M. [2002], *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, Oficyna Wydawnicza SGH, Warszawa.

Harchaoui T.M., Hamdad M. [2000], *The prices of classical recorded music: a hedonic approach*, International Journal of Industrial Organization, **18**, pp. 497–514.

Howkins J. [2001], *The Creative Economy: How People Make Money From Ideas*, Penguin, London.

Hui K-L., Png I. [2003], *Piracy and the legitimate demand for recorded music*, Contributions to Economic Analysis & Policy , **2**.

Lee J., Boatwright P., Kamakura W.A. [2003], *A Bayesian model for prelaunch sales forecasting of recorded music*, Management Science, **49**, pp. 179–196.

Liebowitz S. [2003], *Will MP3 downloads annihilate the record industry? The evidence so far*, University of Texas, Dallas.

Long J.S. [1997]. *Regression models for categorical and limited dependent variables*, Sage Publications, Thousand Oaks.

Long J.S., Freese J. [2005]. *Regression models for categorical dependent variables using Stata*, Stata Press, College Station.

Maddala G.S. [1994]. *Limited-dependent and qualitative variables in econometrics*, Cambridge University Press, Cambridge.

McKelvey R.D., Zavoina W. [1975], *A statistical model for the analysis of ordinal level dependent variables*, Journal of Mathematical Sociology, **4**, pp. 103–120.

Mixon F., Ressler R. [2000], *A Note on Elasticity and Price Dispersions in the Music Recording Industry*, Review of Industrial Organization, **17**, pp. 465–470.

Oberholzer F., Strumpf K. [2007], *The effect of file sharing on record sales: an empirical analysis*, Journal of Political Economy, **115**, pp. 1–42.

Ordanini A. [2006], *Selection models in the music industry: How a prior independent experience may affect chart success*, Journal of Cultural Economics, **30**, pp. 183–200.

Peterson B., Harrell F.E. [1990], *Partial proportional odds models for ordinal response variables*, Applied Statistics , **39**, pp. 205–217.

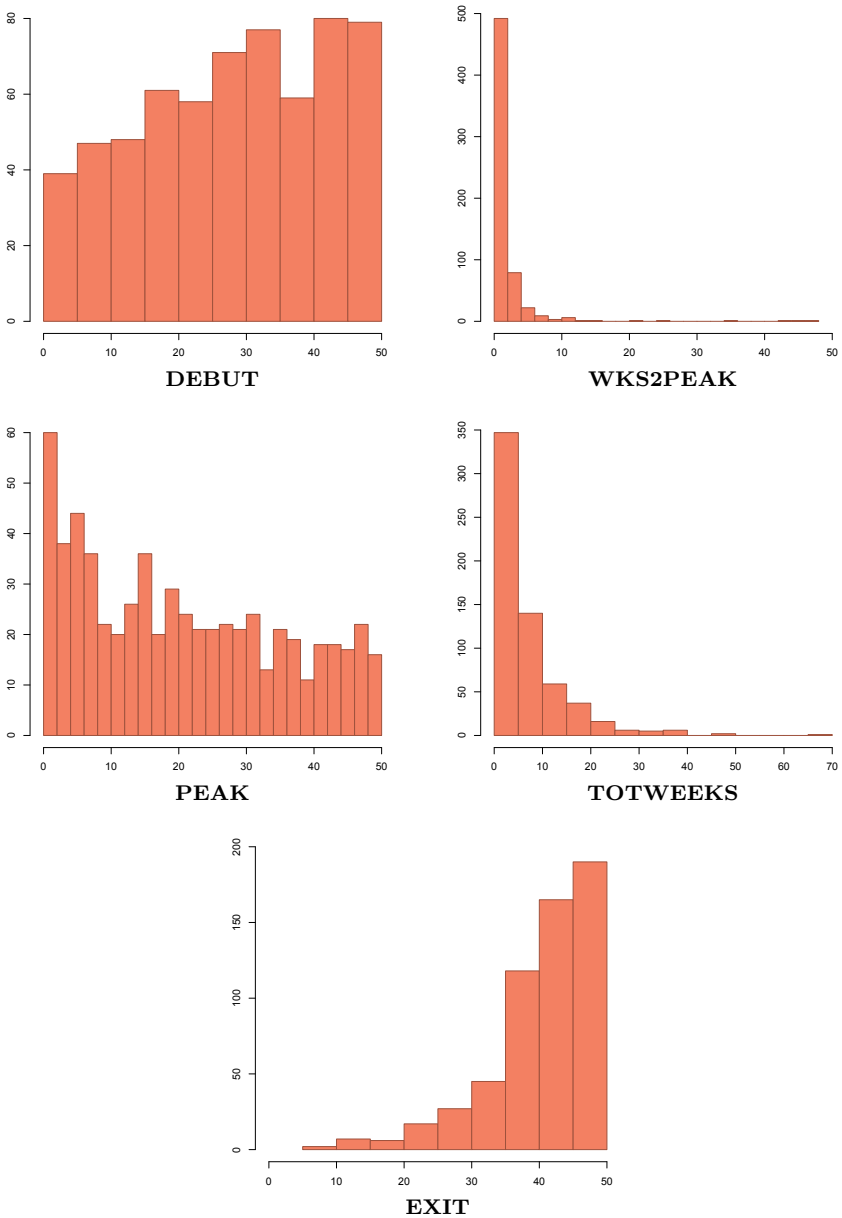
Recording Industry in Numbers 2009, International Federation of the Phonographic Industry, London.

Snell E.J. [1964], *A scaling procedure for ordered categorical data*, Biometrics, **20**, pp. 592–607.

Strobl E., Tucker C. [2000], *The dynamics of chart success in the U.K. pre-recorded popular music industry*, Journal of Cultural Economics, **24**, pp. 113–134.

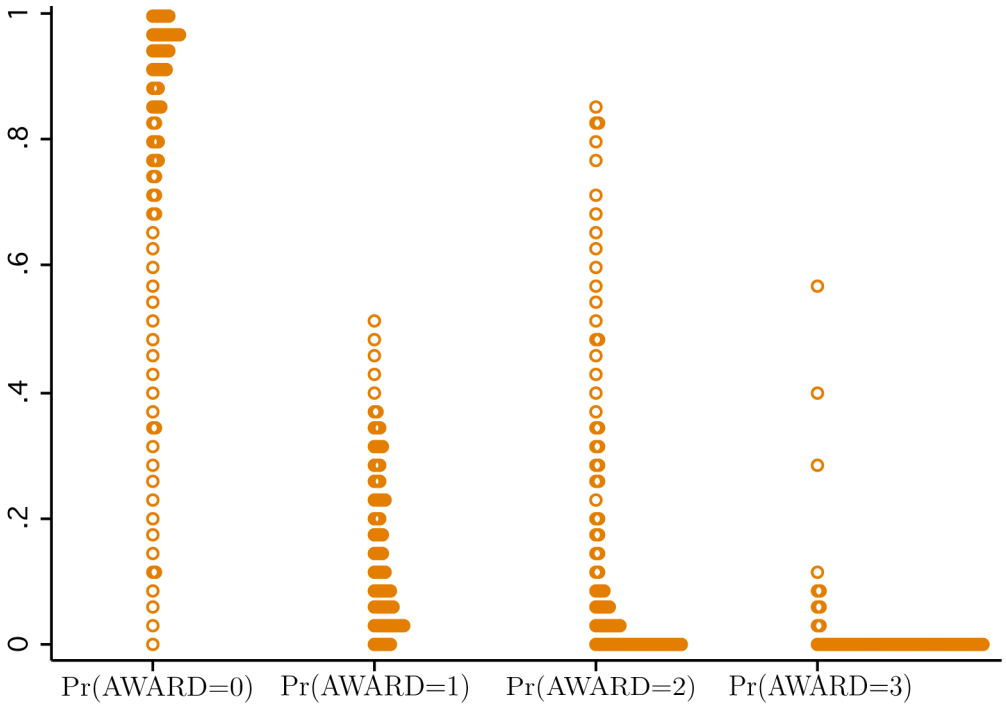
Williams R. [2005], *Generalized ordered logit/partial proportional odds models for ordinal dependent variables*, The Stata Journal , **6**, pp. 58–82.

Figure 1: Histograms for measures representing album's trajectory on the chart.



Source: own computations.

Figure 2: Predicted probabilities for *ZPD*.



Source: own computations.

Table 12: Interdependency matrix (statistically significant coefficients at the 0.05 level typed in bold).

	PRICE	DEBUT	WKS2PEAK	PEAK	TOTWKS	EXIT	MAJOR	PL	VA	OST	CLASSJAZZ	SEASON	**GOLD**	**NUMB1**	**AWARD**
PRICE	1														
DEBUT	**0,09**	1													
WKS2PEAK	-0,01	-0,03	1												
PEAK	0,06	**0,74**	**-0,49**	1											
TOTWKS	-0,03	**-0,49**	**0,65**	**-0,81**	1										
EXIT	0,07	**0,12**	-0,05	**0,11**	0,05	1									
MAJOR	+	-0,03	**0,17**	**-0,09**	**0,24**	**0,13**	1								
PL	—	**-0,09**	-0,03	-0,08	0,03	-0,02	**-0,33**	1							
VA	0	**0,11**	0,03	**0,09**	**-0,08**	-0,05	**0,12**	**-0,24**	1						
OST	0	-0,04	**0,11**	-0,06	**0,13**	0,05	**0,11**	-0,04	**-0,13**	1					
CLASSJAZZ	0	**0,08**	0,04	0,05	0,02	**0,14**	**0,08**	-0,05	0,06	-0,05	1				
SEASON	0	0,04	0,06	0	-0,02	0,04	0,04	**-0,08**	**0,42**	-0,06	-0,05	1			
GOLD	0	**-0,29**	**0,26**	**-0,45**	**0,52**	**0,13**	**0,23**	0,01	-0,02	0,04	**0,16**	0,05	1		
NUMB1	0	**-0,3**	**0,09**	**-0,39**	**0,31**	0,04	0,07	0,06	**-0,1**	-0,01	-0,01	0,01	**0,27**	1	
AWARD	0,03	**-0,3**	**0,28**	**-0,47**	**0,54**	**0,13**	**0,23**	0,01	-0,02	0,05	**0,18**	0,05	**0,98**	**0,32**	1

Source: own computations.