Model Choice Applied to Consumer Preferences

Iain Pardoe, University of Oregon
Catherine A. Durham, Oregon State University
Contact: Iain Pardoe, Charles H. Lundquist College of Business, 1208
University of Oregon, Eugene, OR 97403–1208 (ipardoe@lcbmail.uoregon.edu)

Key Words: Bayesian methodology; Diagnostic; Graphical method; Mixture models; Model assessment; Random effects.

Abstract:

Standard model assessment techniques such as residual plots or Akaike's information criterion can be difficult to use or provide limited insight into model fit when applied in non-standard regression contexts such as random effects or mixture models. This paper considers the application of Bayesian ideas to model choice for a large consumer preference dataset where non-standard models appear to be needed but it is unclear which model is most appropriate. In addition, it is shown that suitably chosen graphical methods can provide insights into which of a set of competing models is most useful for different subsets of the data.

1. Introduction

This article describes a model choice case study of an interesting consumer preference dataset concerning restaurant wines. A variety of count data regression models are plausible for this dataset, but it is not clear which model provides the most useful and parsimonious description of the data. We apply a variety of numerical and graphical diagnostic techniques to the model results, and in graphically presenting these results discover that alternate models appear to be appropriate for different subsets of the data.

Section 2 outlines the application and dataset, while Section 3 gives details of the models considered. The focus of the paper is Section 4 which describes the graphical and numerical model choice techniques applied. Section 5 contains a discussion.

2. Application

The application uses wine sales data for an Oregon restaurant to examine whether wine characteristics, including sensory information available to customers, influence demand. In particular, the analysis uses hedonic quantity models to evaluate the impact of objective (origin, varietal, etc.) and sensory descriptors, and price, on the choice of wine.

The data were collected at a high-end restaurant between the end of April and the beginning of September 1998, a nineteen-week period. The restaurant offers a wide selection of wines (stocking some 1500 bottles), ranging from less expensive to premium reserve wines. It offers wines from a variety of origins and provides a detailed wine menu for its customers. This menu is broken down into white, red and sparkling wines, and further subdivided into varietal and/or origin. It also provides a description of the sensory qualities of the wine along with the brand, vintage, origin, and price. Sensory information usually includes aroma, flavors, and sometimes "mouth feel" (e.g. dry, tannic, smooth). Typical descriptors for aroma and taste include different types of fruits (berry, lemon), flowers (apple, rose), and other food associations such as herbal, honey, and chocolate. There are also numerous, widely accepted terms for mouth feel, concentration, or texture that are not associated with a taste or smell, such as big, creamy, or heavy.

Wine prices are generally based on expert quality assessments, with adjustment for varietal, origin, and market factors. Only in limited circumstances can wine price be said to reflect wine quality, or consumer valuation of those qualities (Combris, Lecocq, and Visser, 2000). Many wines can be substantially over or under priced given relative quality, due to the great variety of wines available, supply variation, and the lack of good information on quality.

There are a number of studies that consider this issue. Oczkowski (1994) uses a hedonic price model to evaluate characteristics influencing Australian wine prices; vintage and a vintage-varietal interaction partly accounted for the endogeneity of quantity supplied. Nerlove (1995) uses an index of quantity purchased and technically assessed wine characteristics to model preferences of Swedish wine purchasers. Finally, Combris, Lecocq, and Visser (1997, 2000) use expert panel jury ratings to evaluate the impact of characteristics on market price for a regional French wine. Only two of the characteristics were significant: whether the wine flavor was concentrated and whether the wine needed extended storage (a positive factor). Further sensory characteristics are significant when modeling the jury rankings themselves. One possible reason for this is that consumers lack perfect information for all characteristics and are thus much more likely to use the "objective" characteristics found on the label (origin, maker, vintage) to make choices. Alternatively, consumer heterogeneity may result in characteristic effects on choice or price offsetting one anther.

Red and white wines usually have different sensory characteristics, their prices have different ranges, and they are selected to go with different foods. Thus, the following variables are considered separately for 47 red and 29 white wines:

- Quantity sold in each of 19 weeks
- Price per bottle as listed on the menu, or four times the per glass price less one dollar for wines available by the glass (the pricing rule for such wines used in the restaurant)
- Low price, an indicator for wines with the lowest price of a particular varietal
- Glass, an indicator for wines sold by the glass
- Origin-Varietal, consisting of

  red: seven indicators for California Cabernet Sauvignon, California Zinfandel, Oregon Pinot Noir, California Other[1], Northwest Other[2], French Red, and Italian Red (relative to the California Merlot base)

  white: five indicators for Oregon Chardonnay, Oregon Pinot Gris, California Other[3], Northwest Other[4], and French White (relative to the California Chardonnay base)
- Ten indicators for sensory characteristics common to reds and whites: Body (full, big, lots of), Finish (long or smooth, etc.), Oak, Rich, Spices
- Fourteen indicators for sensory characteristics unique to reds and whites, consisting of

  red: Currant (black or red), Berry (black, Marion, raspberry), Cherry, Chocolate, Tannic (medium, firm, plenty of), and Vanilla

  white: Buttery, Creamy, Dry, Honey, Melon, Citrus (includes lemon, grapefruit), Tree Fruit (apple, peach, pear), and Tropical Fruit

With sufficient variability in the data, it might have been possible to evaluate origin and varietal effects separately as well as joint effects for specific combinations. However, most of the wines listed for a particular varietal are from a region where that varietal is recognized for good quality. For example, all of the Zinfandels and all but one of the Cabernet Sauvignon selections are from California. Thus, though these could be treated separately, it would be inaccurate to relate a parameter estimate for "Zinfandel" to Zinfandel effects across all origins. Thus, the models presented treat varietals and origins as pairs, though there are a number of wines that are aggregated as more general "others".

The sensory descriptors used in the analysis were the most common ones on the wine menu. Other descriptors applied only to a few wines and thus were excluded from consideration. Non-alcoholic and sparkling wines were excluded because it is reasonable to suppose that the decision to drink such wines excludes consideration of other (alcoholic, non-sparkling) wines. For experienced wine enthusiasts, the combination of vintage, varietal, and origin provides information about the grape quality for a specific wine. According to the restaurant's wine steward, about five percent of the study restaurant's clientele might have some knowledge regarding a good or bad vintage. While model and data limits precluded accurate testing of vintage impacts, it does not seem that this would be very relevant for this population of consumers.

3. Models

Ordinarily, count data such as this would be modeled using log-linear Poisson regression, with the (log) Poisson means dependent on characteristics associated with each wine. However, this data exhibits over-dispersion, with, for example, more zero-counts than a Poisson model allows for: of the 1425 observations, 1000 (70.2 percent) were zero (i.e. no bottles of that wine sold that week), whereas a log-linear Poisson regression model predicts only 67.8 percent zeros. To address this problem, we tried the following models: negative binomial; zero-inflated Poisson (ZIP) (Lambert, 1992); hurdle (Mullahy, 1986); zero-inflated negative binomial (ZINB). To gauge the improvement in fit from accounting for over-dispersion by using such models, we also fit a standard log-linear Poisson regression model.

[1] Includes a Syrah, Petit Syrah, and varietal blend.

[2] Includes Washington or Oregon Cabernets and Merlots.

[3] Includes Fume Blanc, Gewurtztraminer, White Zinfandel and a Sauvignon Blanc/Semillon Blend.

[4] Includes Muller Thurgau, Chenin Blanc, Gewurtztraminer, and Riesling.

3.1 Negative Binomial Model

The standard Poisson regression model can be generalized so that each observation has its own multiplicative random effect in the Poisson mean, with these random effects having a Gamma distribution with mean one and variance $1/\alpha$. Whereas the standard Poisson model restricts the variance to equal the mean, introducing random effects allows the variance to exceed the mean by adjusting the Poisson means to reflect unexpectedly high or low demand. The marginal distribution of the counts (integrating out the random effects) is negative binomial so that the count probabilities are

$$\Pr(Q_i = q) = \frac{\Gamma(q+\alpha)}{q!\,\Gamma(\alpha)} \left(\frac{\alpha}{\alpha+\mu_i}\right)^{\alpha} \left(\frac{\mu_i}{\alpha+\mu_i}\right)^{q}$$

where Q_i denotes the number of bottles of wine sold in a week, and $i = 1, \ldots, n = 1425$ (76 wines sold over a 19-week period, but with one of the red wines replacing another part way through the period).

3.2 ZIP Model

The traditional way in which a ZIP model allows for over-dispersion is to assume that the counts follow a mixture distribution: Poisson(μ_i) with probability p_i or identically zero (i.e. a "structural zero") with probability $1-p_i$, where μ_i is the Poisson mean. The Poisson means are modeled as a function of the wine characteristics, and the structural zero probabilities can either be completely stochastic or can also be modeled as a function of the wine characteristics. We modify this set-up in light of the fact that one of the wine characteristics almost guarantees non-zero (positive) sales: Glass. There were nine wines available by the glass, and of the 171 weekly counts for these wines, only nine were zeros. Such wines were modeled as Poisson(μ_i). Other (non-glass) wines followed the usual ZIP model.

Thus, the count probabilities are

$$\begin{aligned}
\Pr(Q_i = 0) &= \exp(-\mu_i)\,\mathrm{I(Glass{=}1)} \\
&\quad + (1 - p_i + p_i \exp(-\mu_i))\,\mathrm{I(Glass{=}0)} \\
\Pr(Q_i = q) &= (\exp(-\mu_i)\mu_i^q/q!)\,\mathrm{I(Glass{=}1)} \\
&\quad + (p_i \exp(-\mu_i)\mu_i^q/q!)\,\mathrm{I(Glass{=}0)}, \\
&\quad q = 1, 2, \ldots
\end{aligned}$$

3.3 Hurdle Model

The Hurdle model is similar to the ZIP model, except that all the zero counts are structural and the non-zero part of the model is zero-truncated Poisson. By contrast, the ZIP model allows some zeros to arise from the Poisson part of the model. Again, we make a modification for wines available by the glass so that the count probabilities are

$$\begin{aligned}
\Pr(Q_i = 0) &= \exp(-\mu_i)\,\mathrm{I(Glass{=}1)} \\
&\quad + (1 - p_i)\,\mathrm{I(Glass{=}0)} \\
\Pr(Q_i = q) &= (\exp(-\mu_i)\mu_i^q/q!)\,\mathrm{I(Glass{=}1)} \\
&\quad + (p_i \exp(-\mu_i)\mu_i^q/q!(1 - \exp(-\mu_i))) \\
&\quad \mathrm{I(Glass{=}0)}, \quad q = 1, 2, \ldots
\end{aligned}$$

3.4 ZINB Model

The negative binomial model can be zero-inflated similarly, so that the count probabilities are

$$\begin{aligned}
\Pr(Q_i = 0) &= \left(\frac{\alpha}{\alpha+\mu_i}\right)^{\alpha} \mathrm{I(Glass{=}1)} \\
&\quad + \left(1 - p_i + p_i \left(\frac{\alpha}{\alpha+\mu_i}\right)^{\alpha}\right) \mathrm{I(Glass{=}0)} \\
\Pr(Q_i = q) &= \left(\frac{\Gamma(q+\alpha)}{q!\,\Gamma(\alpha)} \left(\frac{\alpha}{\alpha+\mu_i}\right)^{\alpha} \left(\frac{\mu_i}{\alpha+\mu_i}\right)^{q}\right) \\
&\quad \mathrm{I(Glass{=}1)} \\
&\quad + \left(p_i \frac{\Gamma(q+\alpha)}{q!\,\Gamma(\alpha)} \left(\frac{\alpha}{\alpha+\mu_i}\right)^{\alpha} \left(\frac{\mu_i}{\alpha+\mu_i}\right)^{q}\right) \\
&\quad \mathrm{I(Glass{=}0)}, \quad q = 1, 2, \ldots
\end{aligned}$$

3.5 Estimation

The link functions that relate $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ and $\boldsymbol{p} = (p_1, \ldots, p_n)$ to the wine characteristics can be written

$$\begin{aligned}
\log(\boldsymbol{\mu}) &= \boldsymbol{X}_1\boldsymbol{\beta} \\
\mathrm{logit}(\boldsymbol{p}) = \log(\boldsymbol{p}/(1-\boldsymbol{p})) &= \boldsymbol{X}_2\boldsymbol{\eta}
\end{aligned}$$

where $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are covariate matrices with columns corresponding to wine characteristics. The covariate matrices can contain covariates in common, and usually $\boldsymbol{X}_2$ contains a subset of the covariates in $\boldsymbol{X}_1$. For our application, $\boldsymbol{X}_1$ consists of the 44 variables described above, while $\boldsymbol{X}_2$ consists of a constant and Price. Incorporating further covariates in $\boldsymbol{X}_2$ made a negligible improvement in the fit of the models.

We take a Bayesian approach, and hence need to specify prior distributions for α, $\boldsymbol{\beta}$, and $\boldsymbol{\eta}$. With small samples this choice can be critical, but with larger samples (such as in this application) the choice is less crucial, since information in the data heavily outweighs information in the prior. Thus, we give $\log(\alpha)$, $\boldsymbol{\beta}$, and $\boldsymbol{\eta}$ uninformative zero-mean Normal priors with standard deviations of ten. In other words, the only assumption made before doing the analysis is that it is implausible that $\log(\alpha)$, $\boldsymbol{\beta}$, and $\boldsymbol{\eta}$ are more than about plus/minus 20.

Table 1: Goodness of Fit Measures

Model	Fit measures			Predicted probability discrepancies			
	AIC	BIC	DIC	Zero	One	Two +	χ^2
Poisson	2577	2808	2574	−2.4%	4.1%	−1.7%	41.0
Negative binomial	2559	2796	2557	−2.0%	3.8%	−1.8%	32.2
ZIP	2519	2761	2517	0.4%	−0.7%	0.3%	21.9
Hurdle	2588	2830	2576	−0.1%	1.8%	−1.7%	30.2
ZINB	2514	2767	2512	0.4%	−0.5%	0.1%	16.6

We used WinBUGS (Spiegelhalter et al., 2003) software to generate posterior samples for $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$. Four chains of 18,000 iterations each for the various models produced trace plots with a good degree of mixing, and various MCMC convergence diagnostics indicated convergence. In particular, after discarding 9,000 burn-in samples and thinning to retain every 9th sample to reduce autocorrelation (leaving a total of 4,000 posterior samples), the 0.975 quantiles of the corrected scale reduction factor (Brooks and Gelman, 1998, p.438) for the $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ parameters were each 1.3 or less.

4. Model Choice

Table 1 compares the models with respect to Akaike's Information Criterion (AIC, Akaike, 1973), Bayesian (or Schwarz's) Information Criterion (BIC, Schwarz, 1978), and Deviance Information Criterion (DIC, Spiegelhalter et al., 2002), as well as discrepancies between predicted and observed probabilities of zero, one, and two or greater counts, and Pearson χ^2 statistics (based on counts from zero to twenty).

The ZIP and ZINB models both appear to offer substantial improvements in fit over the standard Poisson model, improvements not matched by the negative binomial and hurdle models. On all measures except for BIC, the ZINB model fits a little better than the ZIP model, at the expense of an added degree of complexity.

Pardoe (2001) and Pardoe and Cook (2002) describe "Bayes marginal model plots" (BMMP's) for checking regression model mean functions. These ideas can be extended to check variance functions, which can be very useful in situations such as this where the competing models differ mainly on how they handle the variance.

The basic idea is to first obtain "model-free residuals" using a non-parametric smooth of a plot of Q versus some measurable function of the predictors, h. For this application, setting h equal to the log(fitted means) provides a useful comparison of the models. Calculate the residuals as the difference between Q and the smoothed values, then square these residuals, plot them against h, and use another smooth to produce a model-free variance function estimate.

Similarly, produce model-based variance function estimates for each posterior sample using the iterated expectation relation for conditional variances:

$$\mathrm{Var}(y|h) = \mathrm{E}(\mathrm{Var}(y|x)|h) + \mathrm{Var}(\mathrm{E}(y|x)|h) \qquad (1)$$

In particular, estimate the first term on the right-hand side of (1) by smoothing the (assumed) variance function from the fitted model against h. Estimate the second term on the right-hand side of (1) by smoothing predicted counts from the fitted model against h, calculating the squared differences between these smoothed predicted counts and the predicted counts themselves, and then smooth these squared differences against h. Each model-based variance function estimate is then the sum of these two terms for each posterior sample.

Finally, plot the absolute values of the model-free residuals versus h and superimpose square roots of the model-free and model-based variance function estimates. 100 model-based estimates is usually enough to provide sufficient resolution in the plot without excessive computing overhead. If the model provides a reasonable estimate of the variance function, and if each of the smooths uses the same smoothing parameter so that any estimation bias should approximately cancel, then the model-free smooth should broadly follow the same pattern as the model-based smooths. Any indication that it does not, for example by falling outside the band of model-based smooths for particular ranges of h, suggests that the model can perhaps be improved upon. For this application, the Poisson model clearly fails to track the higher variation in counts at larger values of h = log(fitted means), as can be seen in the upper plot of Figure 1.

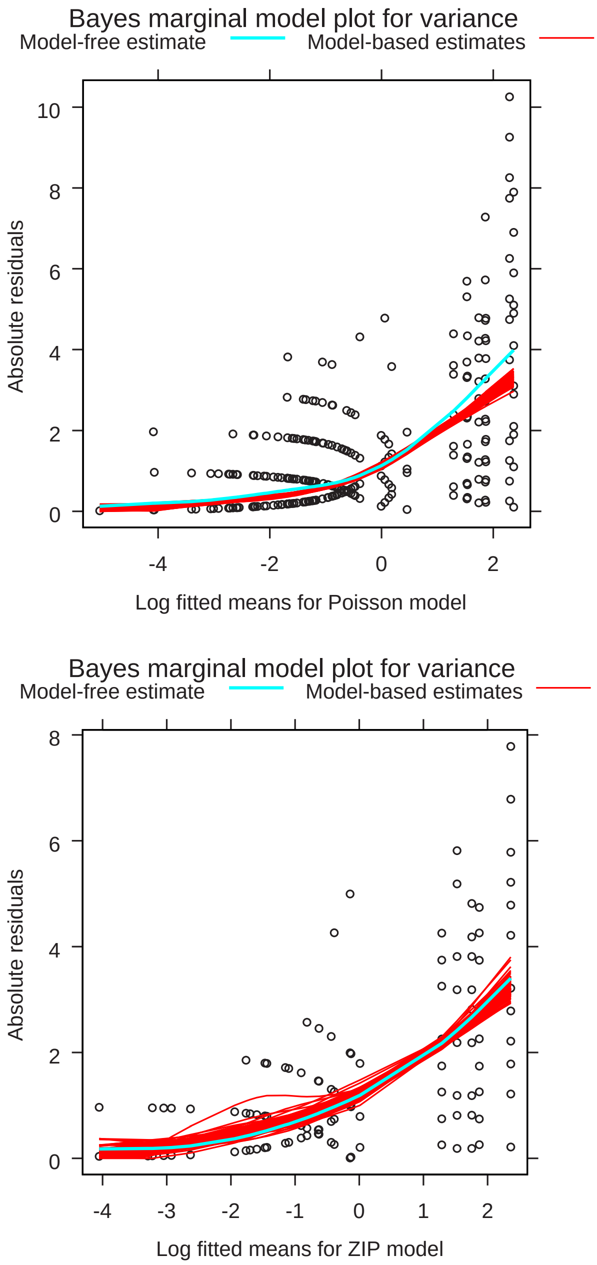

Figure 1: BMMP's for checking the variance functions for all the wines with the Poisson model (upper) and for the white wines with the ZIP model (lower).

BMMPs can be summarized numerically using a Bayes Discrepancy Measure (BDM). First calculate the average squared distance from the model-free smooth to the model-based smooths (call this a, say). Then calculate the average squared distances from each model-based smooth to all the other model-based smooths (call these b_m, $m = 1, \ldots, 100$). Then the BDM is the proportion of b_m's smaller than a; a BDM near zero indicates that the model-free smooth passes close to the center of the model-based smooths, while a BDM near one indicates that the model-free smooth differs greatly from the model-based smooths. Thus low values of BDM indicate a good fit of the variance function, whereas large values indicate a poor fit. For example, the BDM for the upper plot in Figure 1 is 1.0, the largest it can be.

To provide confidence in a particular model, BMMP's should have broadly matching smooths for a variety of h-functions (similar to the requirement for standard residual plots), as well as matching for subsets of the data. We focus on the latter point now, since this dataset naturally comprises two subsets, for white and red wines. The lower plot of Figure 1 shows a BMMP for white wines only with the ZIP model. This shows a clear improvement on the upper plot, and provides additional support for the ZIP model (for white wines at least). The BDM value for this plot is 0.49.

Similarly, the other numerical measures in Table 1 can be calculated separately for white and red wines. A "parallel coordinate" plot provides a useful way to display and compare the results for each of the model choice techniques. Figure 2 displays the relative positions of each model on a common scale for each of the model choice techniques by translating each value to the range (0,1).

By taking into account all of the model choice techniques together, it appears that while the ZIP model shows the most promise for white wines, the ZINB model is preferable for the red wines. Indeed, if a hybrid model which has the ZIP structure for white wines and the ZINB structure for the red wines is fit to this dataset, the resulting goodness of fit measures are AIC = 2507, BIC = 2755, DIC = 2507, $\chi^2 = 17.4$, and BDM = 0.75. Taken together with the results in Table 1, this suggests that such a model may well be the most useful for this dataset.

5. Discussion

Single model fit summaries can sometimes be misleading. Indeed for the wine application presented in this article, the residual deviance for the Poisson model is 1369 on 1381 degrees of freedom; this would ordinarily be suggestive of a reasonable fit. However, to thoroughly investigate whether alternative models might provide a superior fit to a standard model, multiple model choice diagnostics should be used. We considered numerical techniques such as

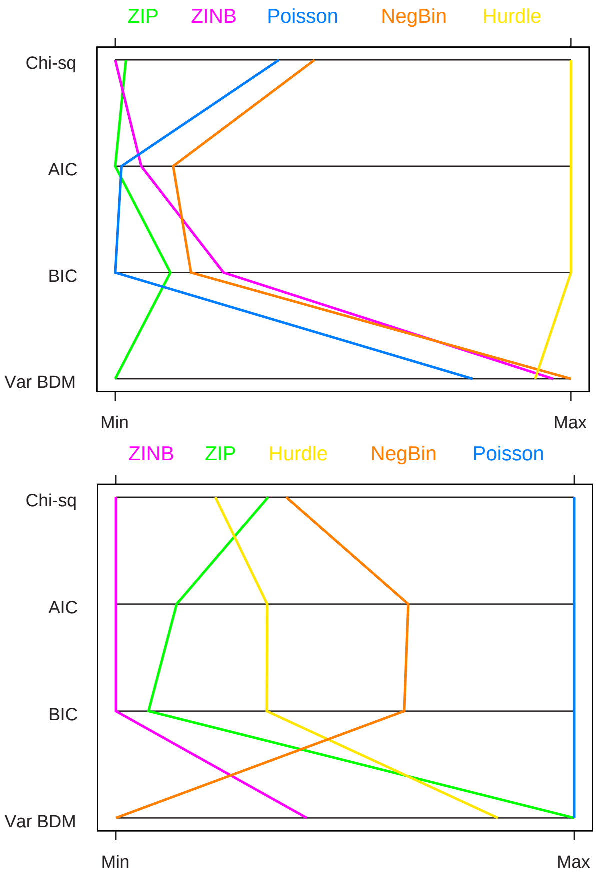

Figure 2: Parallel coordinate plots for goodness of fit measures for white wines (upper) and red wines (lower).

AIC, BIC, DIC, and Pearson's χ^2 statistic, as well as graphical techniques such as Bayes marginal model plots. Taken together, these demonstrated the relatively poor fit of the Poisson model, particularly when compared with zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) models.

Further, simple graphical displays of results can inspire ideas for model improvement. For example, by displaying model fit measures in parallel coordinate plots, we were able to discern particularly well-fitting models for white and red wines separately.

Practical considerations are also important in any model choice exercise. For the wine application, the ZIP and ZINB models offer comparable fits overall, and in fact result in qualitatively similar conclusions about the effects of the wine characteristic covariates. Thus while the hybrid model suggested at the end of Section 4 may offer an ideal solution from a statistical perspective, the best practical solution might be either of the ZIP or ZINB models.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov and F. Csáki (Eds.), *2nd International Symposium on Information Theory*, Budapest. Akadémiai Kiadó.

Brooks, S. P. and A. Gelman (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics 7*, 434–455.

Combris, P., S. Lecocq, and M. Visser (1997). Estimation for a hedonic price equation for Bordeaux wine: Does quality matter? *Economic Journal 107*, 390–402.

Combris, P., S. Lecocq, and M. Visser (2000). Estimation of a hedonic price equation for Burgundy wine. *Applied Economics 32*, 961–967.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics 34*, 1–14.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics 33*, 341–365.

Nerlove, M. (1995). Hedonic price functions and the measurement of preferences: The case of Swedish wine consumers. *European Economic Review 39*, 1697–1716.

Oczkowski, E. (1994). A hedonic price function for Australian premium table wine. *Australian Journal of Agricultural Economics 38*, 93–110.

Pardoe, I. (2001). A Bayesian sampling approach to regression model checking. *Journal of Computational and Graphical Statistics 10*, 617–627.

Pardoe, I. and R. D. Cook (2002). A graphical method for assessing the fit of a logistic regression model. *The American Statistician 56*, 263–272.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*, 461–464.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B (Methodological) 64*, 1–34.

Spiegelhalter, D. J., A. Thomas, N. G. Best, and D. Lunn (2003). *WinBUGS Version 1.4 User Manual*. Cambridge, UK: MRC Biostatistics Unit.