Appendix to “An Empirical Model of Optimal Dynamic Product Launch and Exit Under Demand Uncertainty”

Günter J. Hitsch[1]
Graduate School of Business
University of Chicago

February 2005

[1]Graduate School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637. Phone: (773) 834-7680. Fax: (773) 702-0458. E-mail: guenter.hitsch@chicagogsb.edu

Appendix A: Model Solution

This appendix contains a brief overview of the methods used to solve the firm's dynamic decision process. The presentation is brief and mostly refers to the methods used. A thorough discussion of the various approaches can be found in the references. The model laid out before is a *Markovian decision process* (Rust, 1996). Decisions are made with respect to a discrete (exit) and continuous variable (advertising). A solution to the dynamic programming (DP) problem is described by the value function V, which is defined on the state space $\mathbb{S} = \{0, ..., T\} \times \mathbb{R}^2 \times \mathbb{R}_+$. The value function satisfies the Bellman equation

$$V(\boldsymbol{s}) = \max \left\{ 0, \sup_{\boldsymbol{\alpha}_A \geq 0} \mathbb{E}\left[\pi\left(\boldsymbol{s}, \boldsymbol{\alpha}; \lambda\right) + \beta V(\boldsymbol{s}') | \boldsymbol{s}, \boldsymbol{\alpha} \right] \right\} \qquad \forall \boldsymbol{s} \in \mathbb{S}. \tag{1}$$

The expectation in the Bellman equation is taken with respect to the transition probability of the state vector. The transition density is given by the product of the transition densities for the prior and the goodwill stock,

$$f\left(\boldsymbol{s}_{t+1} | \boldsymbol{s}_t, \boldsymbol{\alpha}_t\right) = f\left(\mu_{t+1} | \boldsymbol{s}_t, \boldsymbol{\alpha}_t\right) \cdot f\left(g_{t+1} | \boldsymbol{s}_t, \boldsymbol{\alpha}_t\right).$$

with $h_{t+1} = \min\left(T, h_t + 1\right)$ and σ_{t+1} given by equation (16) in the main paper. As h_{t+1} and σ_{t+1} evolve deterministically, the integral over the value function has be taken only with respect to μ_{t+1} and g_{t+1} :

$$\mathbb{E}\left[V(\boldsymbol{s}') | \boldsymbol{s}, \boldsymbol{\alpha} \right] = \int_{-\infty}^{\infty} \int_{0}^{\infty} V(\boldsymbol{s}') f\left(\boldsymbol{s}' | \boldsymbol{s}, \boldsymbol{\alpha}\right) dg' d\mu'.$$

The expectation of current-period profit function is found by integrating with respect to the prior on λ :

$$\mathbb{E}\left[\pi\left(\boldsymbol{s}, \boldsymbol{\alpha}; \lambda\right) | \boldsymbol{s}, \boldsymbol{\alpha} \right] = \int_{-\infty}^{\infty} \pi\left(\boldsymbol{s}, \boldsymbol{\alpha}; \lambda\right) p\left(\lambda; \boldsymbol{s}\right) d\lambda.$$

The right-hand side of the Bellman equation (1) defines a map from the complete vector-space of bounded, real-valued functions on $\mathbb{S}$ into itself. This map is a contraction-mapping; hence, the dynamic programming problem has a unique solution. This argument is standard (Bertsekas 1995) under the assumptions made about the payoff function π and the transition densities. Once a solution of the Bellman equation is found, the optimal policy $\boldsymbol{\alpha}(\boldsymbol{s}) = (E, A)$ is implicitly defined through the maximization of the right-hand side of the Bellman equation. In each period the firm compares the value from continuing the product to the exit value, which is 0. If the expected value from staying in the market is negative, the product exits, i.e. $E = 0$. Otherwise, $E = 1$, and the firm chooses an advertising level which maximizes the sum of the expected present discounted value of the product. The optimal policy depends only on the current realization of the state vector. Hence, the firm cannot increase the expected PDV of profits by basing its decisions on the history of realized state variables until t. We solved the dynamic programming problem using a discrete policy iteration (DPI) algorithm, as described in Benítez-Silva et. al. (2003).[1] As part of this

[1] See Judd (1998), Rust (1996), and Benítez-Silva et. al. (2003) for overviews of solution methods to dynamic programming problems. The latter paper contains further details on the numerical methods used to solve this particular problem, and a discussion of alternative solution approaches.

method, the continuous dimensions of the state space were discretized, and the value function V was represented on the resulting grid points. All expectations were calculated using quadrature methods. The DPI algorithm starts with an initial guess for the policy function, $\boldsymbol{\alpha}_0$. The value function, and the policy rule are then updated using the *Bellman operator* implicitly defined by equation (1), until both the value and policy functions converge with respect to some metric, i.e. until $\|V_{i+1} - V_i\| < \varepsilon_V$ and $\|\boldsymbol{\alpha}_{i+1} - \boldsymbol{\alpha}_i\| < \varepsilon_\alpha$.[2] The DP can be numerically solved given current (2004) standard computing technology, and the solution time is fast enough such that the solution can be used in the estimation procedure. For example, a version of the model with a discretized state space of 35,280 points was solved in about five seconds on a computer equipped with a dual-processor Pentium Xeon 2.2 GHz CPU.[3] A relatively short solution time is necessary to make the ML estimator described in the next section feasible, because each evaluation of the log-likelihood function requires us to solve the DP J (the number of products) times. We typically found that the DP was 'well behaved', i.e. convergence occurred within a few steps of the DPI algorithm.

Appendix B: Derivation of the Log-Likelihood Function

Definitions We observe a cross section of products i, and data $\boldsymbol{y}_{it} = (h_{it}, A_{it}, \boldsymbol{\xi}_{it}, E_{i,t+1})$. The data history until period t is denoted by $\boldsymbol{y}_i^t = (\boldsymbol{y}_{i0}, ..., \boldsymbol{y}_{it})$. $\boldsymbol{y}_i$ denotes the whole sample history for product i.

Conditional densities We attempt to construct the probability density of the sample vector $\boldsymbol{y}_i$ as the product of the conditional densities $f\left(\boldsymbol{y}_{it}|\boldsymbol{y}_i^{t-1}, p_{i0}; \boldsymbol{\theta}\right)$. As will become clear in the discussion below, it is simpler to start with a slightly different approach. In this approach, we initially construct the likelihood of $\boldsymbol{z}_i = (\boldsymbol{y}_i, \boldsymbol{g}_i)$, i.e. we start from the assumption that the goodwill stocks are observed. This assumption allows us to construct the probability density of $\boldsymbol{z}_i$,

$$f\left(\boldsymbol{z}_i|p_{i0};\boldsymbol{\theta}\right) = \left[\prod_{t=1}^{T} f\left(\boldsymbol{z}_{it}|\boldsymbol{z}_i^{t-1}, p_{i0}; \boldsymbol{\theta}\right)\right] f\left(\boldsymbol{z}_{0t}|p_{i0};\boldsymbol{\theta}\right). \quad (2)$$

p_{i0} is the initial condition–the prior belief on the product 'quality' parameter λ at the beginning of the sample period. Once we have accomplished this task, we find the marginal distribution of $\boldsymbol{y}_i$ by integrating over $\boldsymbol{g}_i$. At the heart of the construction of the conditional densities is the reconstruction of the unobserved history of state vectors $\boldsymbol{s}_{it}$, which is possible given the initial condition p_{i0} and the data vectors $\boldsymbol{z}_{it}$.

[2] For all parameter values, we calculated the model solution using different numbers of grid points and quadrature nodes in order to achieve a robust model solution.

[3] The code was written in C++, and compiled using the highly efficient Intel C++ compiler. Alternatively, the coding could be performed in a matrix programming language such as MATLAB; however, the solution time would increase significantly.

Reconstruction of the state vectors Assume that we already know the state vector $\boldsymbol{s}_t$, and observe the 'data' $\boldsymbol{z}_i^t$. The M-vector of price-adjusted mean utilities,

$$\begin{aligned}&\boldsymbol{\xi}_t = \varphi\left(\boldsymbol{s}_t, A_t\right) + \boldsymbol{\varepsilon}_t + \boldsymbol{\nu}_t,\\ &\text{where} \quad \varphi_i\left(\boldsymbol{s}_t, A_t\right) = \lambda_i + \psi(g_{it}, A_{it}) + \tau\left(h_{it}\right),\end{aligned}$$

is part of our data. From this equation we can calculate $\boldsymbol{\varepsilon}_t + \boldsymbol{\nu}_t$,the signal used by the firm to update its prior on the product quality λ. Hence, the prior p_{t+1} is uniquely determined. h_t evolves deterministically, and thus only g_{t+1} needs to be known to determine the whole state vector $\boldsymbol{s}_{t+1}$. We can summarize this discussion in the following lemma:

Lemma. *Given knowledge of the initial condition p_{i0} and the data history $\boldsymbol{z}_i^{t-1}$, the state vector $\boldsymbol{s}_{i,t-1}$ and the prior p_{it} are uniquely determined.*

Restrictions on the goodwill process The advertising policy places some restrictions on the data that can be observed, in particular on the goodwill levels. To see this, remember some properties of the advertising policy function $\boldsymbol{\alpha}_A\left(\boldsymbol{s}_t; \boldsymbol{\theta}\right)$. Fix all components of the state vector apart from the goodwill stock g, and denote the resulting policy, which is now only a function of goodwill, by $\chi_t(g) = \chi_t(g; \boldsymbol{\theta})$. If predicted advertising is positive at $g = 0$, i.e. $\chi_t(0) > 0$, the advertising policy function is strictly decreasing in goodwill g over some finite interval. As the marginal revenue of advertising goes to zero for large g, there is some g' such that $\chi_t(g) = 0$ for all $g \geq g'$. Let $\bar{g}$ be the smallest g' for which the previous condition is true, i.e. $\bar{g} = \inf\left\{g' : \chi(g) = 0 \text{ for all } g \geq g'\right\}$. Hence, the interval on which advertising is strictly decreasing in g is given by $[0, \bar{g})$. This property of the advertising policy has several implications for the goodwill process that can be observed, i.e. is feasible given the model restrictions. First, if $0 < A_t \leq \chi_t(0)$, only the goodwill stock $g_t = \chi_t^{-1}(A_t)$ can have generated observed advertising. If $A_t = 0$, we cannot invert the advertising policy to reconstruct the goodwill level. We know, however, that $g_t \geq \bar{g}_t$. Finally, if $A_t > \chi_t(0)$, observed advertising is not explained at the current initial condition and parameter vector. This case does not require further attention, because it implies a joint likelihood of 0.

Conditional distribution of (A_{it}, g_{it}) Remember that goodwill evolves in the form $g_t = \rho_t g_{t-1}^a$, where the stochastic depreciation factor ρ_t has the lognormal distribution $\log(\rho_t) \sim N(\mu_\rho, \sigma_\rho^2)$ with associated density f_ρ. To derive the conditional density of A and g, we first distinguish between the cases where advertising is positive or zero. First, if $0 < A_t \leq \chi_t(0)$,we know from the discussion above that $g_t = \chi_t^{-1}(A_t)$. For any $A \in (0, \chi_t(0)]$,

$$\Pr\left\{A_t \leq A | 0 < A_t \leq \chi_t(0), \boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\} = \frac{\Pr\left\{A_t \leq A \text{ and } 0 < A_t \leq \chi_t(0) | \boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\}}{\Pr\left\{0 < A_t \leq \chi_t(0) | \boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\}}.$$

As $A \leq \chi_t(0)$, the numerator is given by

$$\begin{aligned}
\Pr\left\{\ 0 < A_t \leq A|\boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\} &= \Pr\left\{\chi_t^{-1}(A) \leq \rho_t g_{t-1}^a < \bar{g}_t|\boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\} \\
&= \Pr\left\{\frac{\chi_t^{-1}(A)}{g_{t-1}^a} \leq \rho_t < \frac{\bar{g}_t}{g_{t-1}^a}|\boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\} \\
&= \int_{\chi_t^{-1}(A)/g_{t-1}^a}^{\bar{g}_t/g_{t-1}^a} f_\rho\left(\rho\right) d\rho.
\end{aligned}$$

To simplify notation, define $\mathfrak{P}_t \equiv \Pr\left\{0 < A_t \leq \chi_t(0)|\boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\}$. We can now derive the conditional density of A on $(0, \chi_t(0)]$:

$$\begin{aligned}
f\left(A|0 < A_t \leq \chi_t(0), \boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right) &= \frac{d}{dA}\Pr\left\{A_t \leq A|0 < A_t \leq \chi_t(0), \boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\} \\
&= \frac{1}{\mathfrak{P}_t}(-1) f_\rho\left(\frac{\chi_t^{-1}(A)}{g_{t-1}^a}\right)\frac{d\chi_t^{-1}(A)}{dA}\frac{1}{g_{t-1}^a} \\
&= \frac{1}{\mathfrak{P}_t} f_\rho\left(\frac{\chi_t^{-1}(A)}{g_{t-1}^a}\right)\left|\frac{d\chi_t^{-1}(A)}{dA}\right|\frac{1}{g_{t-1}^a}
\end{aligned}$$

The joint density of A and g is then given by

$$\begin{aligned}
&f\left(A, g|0 < A_t \leq \chi_t(0), \boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right) = \\
&\qquad = \begin{cases} \frac{1}{\mathfrak{P}_t} f_\rho\left(\frac{\chi_t^{-1}(A)}{g_{t-1}^a}\right)\left|\frac{d\chi_t^{-1}(A)}{dA}\right|\frac{1}{g_{t-1}^a} & \text{if } g = \chi_t^{-1}(A) \text{ and } A \in (0, \chi_t(0)], \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

Second, if $A_t = 0$, our focus is on finding the conditional distribution of the current goodwill stock:

$$\Pr\left\{g_t \leq g|A_t = 0, \boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\} = \frac{\Pr\left\{g_t \leq g \text{ and } A_t = 0|\boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\}}{\Pr\left\{A_t = 0|\boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\}}.$$

Define $\breve{g}_t = \min\left\{g : \chi_t\left(g'\right) = 0 \text{ and } \boldsymbol{\alpha}_E(\boldsymbol{s}_t; \boldsymbol{\theta})|_{g_t = g'} = 0 \text{ for all } g' \geq g\right\}$. Furthermore, decompose the data vector $\boldsymbol{z}^{t-1}$ as follows: $\boldsymbol{z}^{t-1} = \left(\tilde{\boldsymbol{z}}^{t-1}, E_t\right)$. Then

$$\begin{aligned}
\Pr\left\{g_t \leq g \text{ and } A_t = 0|\boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\} &= \Pr\left\{g_t \leq g \text{ and } A_t = 0|E_t = 0, \tilde{\boldsymbol{z}}^{t-1}, p_0; \boldsymbol{\theta}\right\} \\
&= \frac{\Pr\left\{g_t \leq g \text{ and } A_t = 0 \text{ and } E_t = 0|\tilde{\boldsymbol{z}}^{t-1}, p_0; \boldsymbol{\theta}\right\}}{\Pr\left\{E_t = 0|\tilde{\boldsymbol{z}}^{t-1}, p_0; \boldsymbol{\theta}\right\}}.
\end{aligned}$$

The numerator can be calculated from

$$\begin{aligned}\Pr\left\{g_t \le g \text{ and } A_t = 0 \text{ and } E_t = 0|\tilde{\boldsymbol{z}}^{t-1}, p_0; \boldsymbol{\theta}\right\} &= \Pr\left\{\breve{g}_t \le g_t \le g|\tilde{\boldsymbol{z}}^{t-1}, p_0; \boldsymbol{\theta}\right\} \\ &= \Pr\left\{\breve{g}_t \le \rho_t g_{t-1}^a \le g|\tilde{\boldsymbol{z}}^{t-1}, p_0; \boldsymbol{\theta}\right\} \\ &= \Pr\left\{\frac{\breve{g}_t}{g_{t-1}^a} \le \rho_t \le \frac{g}{g_{t-1}^a}|\tilde{\boldsymbol{z}}^{t-1}, p_0; \boldsymbol{\theta}\right\} \\ &= \int_{\breve{g}_t/g_{t-1}^a}^{g/g_{t-1}^a} f_\rho\left(\rho\right) d\rho.\end{aligned}$$

Furthermore, using $\tilde{g}_t = \min\left\{g : \boldsymbol{\alpha}_E(\boldsymbol{s}_t; \boldsymbol{\theta})|_{g_t = g'} = 0 \text{ for all } g' \ge g\right\}$, we find that

$$\mathfrak{Q}_t \equiv \Pr\left\{E_t = 0|\boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\} = \int_{\breve{g}_t/g_{t-1}^a}^{\infty} f_\rho\left(\rho\right) d\rho.$$

From the definition above, $\Pr\left\{A_t = 0|\boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\} = 1 - \mathfrak{P}_t$. We can now derive the conditional density of g:

$$\begin{aligned}f\left(g|A_t = 0, \boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right) &= \frac{d}{dg}\Pr\left\{g_t \le g|A_t = 0, \boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\} \\ &= \frac{1}{\mathfrak{P}_t}\frac{d}{dg}\Pr\left\{g_t \le g \text{ and } A_t = 0|\boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\} \\ &= \frac{1}{\mathfrak{P}_t\mathfrak{Q}_t} f_\rho\left(\frac{g}{g_{t-1}^a}\right)\frac{1}{g_{t-1}^a}.\end{aligned}$$

In summary, the joint density of g and A is

$$f\left(A, g|A_t = 0, \boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right) = \begin{cases} \frac{1}{\mathfrak{P}_t\mathfrak{Q}_t} f_\rho\left(\frac{g}{g_{t-1}^a}\right)\frac{1}{g_{t-1}^a} & \text{if } g \ge \breve{g}_t \text{ and } A = 0, \\ 0 & \text{otherwise.}\end{cases}$$

Note that $A_t > \chi_t(0)$ is inconsistent with the model, and implies a zero joint density. The densities conditional on A_t allows us to find the unconditional (on A) joint density of A and g :

$$\begin{aligned}f\left(A, g|\boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right) = f\left(A, g|0 < A_t \le \chi_t(0), \boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right)\Pr\left\{0 < A_t \le \chi_t(0)|\boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\} \\ + f\left(A, g|A_t = 0, \boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right)\Pr\left\{A_t = 0|\boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right\}.\end{aligned}$$

In summary,

$$f\left(A, g|\boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\right) = \begin{cases} f_\rho\left(\frac{\chi_t^{-1}(A)}{g_{t-1}^a}\right)\left|\frac{d\chi_t^{-1}(A)}{dA}\right|\frac{1}{g_{t-1}^a} & \text{if } g = \chi_t^{-1}(A) \text{ and } A \in (0, \chi_t(0)], \\ \frac{1}{\mathfrak{Q}_t} f_\rho\left(\frac{g}{g_{t-1}^a}\right)\frac{1}{g_{t-1}^a} & \text{if } g \ge \breve{g}_t \text{ and } A = 0, \\ 0 & \text{otherwise.}\end{cases}$$

Conditional distribution of $\boldsymbol{\xi}_{it}$ Once again, the price-adjusted mean utility can be expressed as

$$\boldsymbol{\xi}_t = \varphi(\boldsymbol{s}_t, A_t) + \boldsymbol{\varepsilon}_t + \boldsymbol{\nu}_t.$$

Hence, conditional on $(A_t, g_t, \boldsymbol{z}^{t-1}, p_0)$, the vector $\boldsymbol{\xi}_t = (\xi_{t1}, ..., \xi_{tM})$ has the multivariate normal distribution $\boldsymbol{\xi} \sim N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma})$, with mean

$$\boldsymbol{\mu}_t = \begin{bmatrix} \varphi(\boldsymbol{s}_t, A_t) \\ \vdots \\ \varphi(\boldsymbol{s}_t, A_t) \end{bmatrix}$$

and covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_\nu^2 & \cdots & \sigma_\varepsilon^2 \\ \vdots & \ddots & \vdots \\ \sigma_\varepsilon^2 & \cdots & \sigma_\varepsilon^2 + \sigma_\nu^2 \end{bmatrix}.$$

The conditional density of $\boldsymbol{\xi}$ is given by a multivariate normal probability,

$$f(\boldsymbol{\xi}|A_t, g_t, \boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}) = f_{\boldsymbol{\xi}}(\boldsymbol{\xi}; \boldsymbol{\mu}_t, \boldsymbol{\Sigma}).$$

Conditional distribution of E_{t+1} First, note that the data $(\boldsymbol{\xi}_t, A_t, g_t, \boldsymbol{z}^{t-1}, p_0) = (\tilde{\boldsymbol{z}}^t, p_0)$ determine next period's prior p_{t+1}. Remember the definitions from above:
$\{\tilde{g}_t = \min g : \boldsymbol{\alpha}_E(\boldsymbol{s}_t; \boldsymbol{\theta})|_{g_t = g'} = 0 \text{ for all } g' \geq g\}$, and

$$\mathfrak{Q}_{t+1} \equiv \Pr\{E_t = 0|\boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}\} = \int_{\tilde{g}_t/g_{t-1}^a}^{\infty} f_\rho(\rho)\, d\rho.$$

Hence,

$$f(E_{t+1}|\boldsymbol{\xi}_t, A_t, g_t, \boldsymbol{z}^{t-1}, p_0; \boldsymbol{\theta}) = \begin{cases} \mathfrak{Q}_{t+1} & \text{if } E_{t+1} = 0, \\ 1 - \mathfrak{Q}_{t+1} & \text{if } E_{t+1} = 1. \end{cases}$$

The joint distributions The previous results allow us to calculate the joint density describing the transition process of $\boldsymbol{z}_{it}$:

$$f(\boldsymbol{z}^t|\boldsymbol{z}^{t-1}, p_{i0}; \boldsymbol{\theta}) = f(E_{t+1}|\boldsymbol{\xi}_t, A_t, g_t, \boldsymbol{z}^{t-1}, p_{i0}; \boldsymbol{\theta}) \cdot f(\boldsymbol{\xi}|A_t, g_t, \boldsymbol{z}^{t-1}, p_{i0}; \boldsymbol{\theta}) \cdot f(A, g|\boldsymbol{z}^{t-1}, p_{i0}; \boldsymbol{\theta}).$$

At this stage, we have determined the joint conditional distribution of $\boldsymbol{z}_i$,given in equation (2) by the formula for $f(\boldsymbol{z}_i|p_{i0}; \boldsymbol{\theta})$. Remember that $\boldsymbol{z}_i = (\boldsymbol{y}_i, \boldsymbol{g}_i)$ is only partially observed. In order to calculate the marginal distribution of $\boldsymbol{y}_i$, we need to integrate over the sequence of goodwill levels $\boldsymbol{g}_i$. In those periods where advertising is positive, goodwill has a mass point at $g_t = \chi_t^{-1}(A_t)$. If, however, $A_t = 0$, we have to integrate over $[\bar{g}_t, \infty)$. Let $\tau(i, 1), ..., \tau(i, L_i)$ denote the $L_i \leq T$ time periods in which $A_{it} = 0$. Then

$$f\left(\boldsymbol{y}_i | p_{i0}; \boldsymbol{\theta}\right) = \int_{\bar{g}_{\tau(i,1)}}^{\infty} \cdots \int_{\bar{g}_{\tau(i,L_i)}}^{\infty} f\left(\boldsymbol{y}_i, \boldsymbol{g} | p_{i0}; \boldsymbol{\theta}\right) dg_{\tau(i,L_i)} \cdots dg_{\tau(i,1)}. \tag{3}$$

The prior p_{i0} is a normal probability distribution with mean μ_{i0} and standard deviation σ_0. μ_{i0} is drawn from $N(\lambda_i, \sigma_0)$. The product quality parameter, λ_i, and the initial standard deviation of the belief, σ_0, are part of $\boldsymbol{\theta}$. The log-likelihood for product i can then be obtained by integrating (3) with respect to normal probability density $f\left(\mu_{i0}; \boldsymbol{\theta}\right)$. Finally, the log-likelihood function of the whole sample is

$$\begin{aligned} l(\boldsymbol{y}; \boldsymbol{\theta}) &= \sum_{i=1}^{N} \log \left(\int_{-\infty}^{\infty} f\left(\boldsymbol{y}_i | \mu_{i0}; \boldsymbol{\theta}\right) f\left(\mu_{i0}; \boldsymbol{\theta}\right) d\mu_{i0} \right) \\ &= \sum_{i=1}^{N} \log \left(\int_{-\infty}^{\infty} \left[\int_{\bar{g}_{\tau(i,1)}}^{\infty} \cdots \int_{\bar{g}_{\tau(i,L_i)}}^{\infty} f\left(\boldsymbol{y}_i, \boldsymbol{g} | \mu_{i0}; \boldsymbol{\theta}\right) dg_{\tau(i,L_i)} \cdots dg_{\tau(i,1)} \right] f\left(\mu_{i0}; \boldsymbol{\theta}\right) d\mu_{i0} \right). \end{aligned}$$

Standard errors The estimation of the asymptotic covariance matrix of the ML estimator is standard. A possible complication arises due to the two step procedure. However, because the first step estimator converges at a faster rate than the second step estimator, the asymptotic distribution of the second step estimator is not influenced by the first step estimator. Our discussion closely follows Wooldridge (2001); please see his book, in particular chapter 12.4, for a complete argument. Let $\boldsymbol{\theta}^{(1)}$ denote the first stage estimator, and $\boldsymbol{\theta}^{(2)}$ denote the second stage estimator. Let $\boldsymbol{s}_i(\boldsymbol{\theta}^{(2)}; \boldsymbol{\theta}^{(1)})$ be the score of the log-likelihood of observation i (note that we previously used the same symbol, $\boldsymbol{s}$, for the state vector). Furthermore, define

$$\boldsymbol{F}_0 = \mathbb{E}\left[\nabla_{\boldsymbol{\theta}^{(1)}} \boldsymbol{s}(\boldsymbol{y}, \boldsymbol{\theta}^{(2)}; \boldsymbol{\theta}^{(1)})\right].$$

$\boldsymbol{F}_0$ is the expected derivative of the score with respect to the first stage estimator. A mean value expansion (Wooldridge 2001, p. 355) gives

$$N^{-1/2} \sum_{i=1}^{N} \boldsymbol{s}_i(\boldsymbol{\theta}_0^{(2)}; \boldsymbol{\theta}^{(1)}) = N^{-1/2} \sum_{i=1}^{N} \boldsymbol{s}_i(\boldsymbol{\theta}_0^{(2)}; \boldsymbol{\theta}^{(1)}) + \boldsymbol{F}_0 \sqrt{N} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}_0^{(1)}\right) + o_p(1). \tag{4}$$

Note that the first stage estimates on which the second step estimator depends, i.e. the price coefficient, interaction terms, market dummies, and time dummies, are $\sqrt{N \cdot M \cdot T}$ consistent, where N is the number of products, M is the number of geographic markets, and T is the number of time periods. The ML estimator, on the other hand, is only $\sqrt{N}$ consistent. Hence,

$$\boldsymbol{F}_0 \sqrt{N} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}_0^{(1)}\right) = o_p(1). \tag{5}$$

Substituting (5) into (4), we see that the influence of the first step estimator on the score, and hence the asymptotic distribution of the second step estimator vanishes asymptotically. Following standard asymptotic arguments (Wooldridge 2001, p. 355), the estimate of the covariance matrix of the second step estimator can then be obtained by plugging the first step estimator into the score or Hessian of the second step estimator.

A note on the maximization algorithm We used the Nelder-Meade downhill simplex algorithm to maximize the log-likelihood function.[4] The restrictions on the observed data imply that not all sample realizations can be explained at some particular parameter vector $\boldsymbol{\theta}$. I.e., for any $\boldsymbol{y}$, there exists a $\boldsymbol{\theta}$ such that $l(\boldsymbol{y};\boldsymbol{\theta}) = -\infty$. This is important for the actual algorithm used to maximize the log-likelihood. For most starting values we found that $l = -\infty$; to overcome this issue we constructed a measure indicating how many conditional densities $f(\boldsymbol{z}_i|p_{i0};\boldsymbol{\theta})$ were 0, and based on this measure we added a 'punishment' value to the log-likelihood. This modified maximization algorithm would always end up in the region of the parameter space where the log-likelihood function was well-defined.

Appendix C: Data Description

The main data set used was collected by IRI, a Chicago-based marketing research company.[5] The data are based on a sample of supermarkets in various U.S. cities. The supermarkets sampled account for more than 80% of all grocery sales in the U.S. IRI makes the sample as representative as possible, because the data are mainly used by the cereal manufacturers to evaluate the sales performance of their products. The data are quarterly, and cover the period between the first quarter of 1988 and the last quarter of 1992. The data set contains both aggregate, national level observations, and a cross-sectional dimension of between 47 and 65 cities in different regions of the U.S.[6] The sales of each cereal brand are aggregated across different box sizes.[7] We observe volume sales, measured in pounds, and dollars sales. The price of a cereal is obtained by dividing dollar sales by volume sales, i.e. prices are measured in terms of $/lb. All prices are deflated using the consumer price index. We define the total market size as the potential volume of breakfast cereal which can be consumed in the whole U.S. or, in the case of a regional market, a given city. We assume that each person in a market can consume one pound per week, and then calculate the total market size as the product of the population size in a city (or the whole U.S.) and the number of weeks in a quarter. The market share of a cereal is then defined as the ratio of its volume sales relative to the whole market size. The market share of the outside good is calculated as the difference between the sum of the market shares of all cereals and 1. The sales data in supermarkets are adjusted for price promotions, and retailer coupons. The redemption of manufacturer coupons, however, is not reflected in the data. Therefore, the measured prices overestimate the actual price paid by consumers for a cereal. Whether this induces a bias in the estimates of the price elasticities is hard to assess; the work of Nevo and Wolfram (2002) suggests that maybe price elasticities could

[4] See Press, Teukolsky, Vetterling, and Flannery (2002) for a detailed description.

[5] Prof. Ron Cotterill, the director of the Food Marketing Policy Center at the University of Connecticut, kindly provided me with the data.

[6] The regional coverage of the database is extended from 47 cities in 1988 to 65 cities in 1992.

[7] Most cereals are sold in different box sizes.

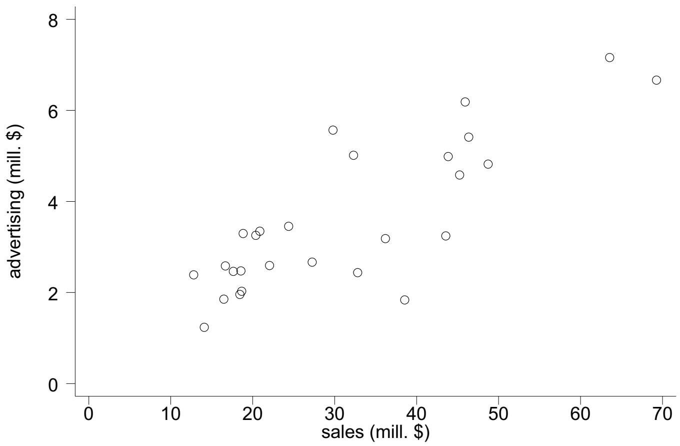

Figure 1: Advertising and sales

be somewhat overestimated.[8] Profits will be overestimated by the presence of coupons, which may cause the model to predict product exit incorrectly. The advertising data used come from the Leading National Advertisers (LNA) data set, and cover national advertising spending in 10 different media[9]. Advertising is measured in dollars, and is available by brand and quarter. Table 1 lists all cereal products included in the empirical analysis.

Some stylized facts on cereal advertising Figure depicts the relationship between advertising and sales for the 27 largest cereal brands.[10] For this graph, advertising expenditures and sales were averaged over the 1988–1992 period. It is evident that there is a strong positive relationship between the two variables; larger products are advertised more. The average advertising/sales ratio for these products is 0.12. Does brand level advertising come out of a constant firm level advertising budget? In particular, does advertising for a new brand reduce the advertising expenditures for an existing brand? Some evidence against this possibility is provided in figure , which shows firm level advertising for the four largest cereal producers. Firm level advertising is less variable than brand level advertising, but not constant. In particular, Kellogg's total advertising expenditures decreased between 1988 and 1992, possibly related to the smaller number of new products introduced and advertised in the later years.

Appendix D: Additional Estimation Results

This section contains some additional sample statistics and estimation results that were not presented in the estimation section. Table 2 shows the sample statistics for the demographic variables

[8]Nevo and Wolfram (2002) find that the shelf prices for a cereal brand are usually lower during periods when coupons are available.

[9]These media are magazines, Sunday magazines, newspapers, outdoor, network television, spot television, syndicated television, cable TV networks, network radio, and national spot radio.

[10]This relationship still holds if a larger number of cereals is included.

Table 1: List of Breakfast Cereals Used in the Empirical Analysis

General Mills	Kellogg	Post	Quaker	Ralston	Nabisco
Apple Cinnamon Cheerios*,e	All Bran	40% Bran Flakes	100% Natural	Chex	Shredded Wheat
Cheerios*	Apple Jacks	Alpha Bits	Cap'n Crunch*	Oat Bran Options*,s	
Cinnamon Toast Crunch*	Bran Flakes*	Cocoa Pebbles	Life		
Clusters	Common Sense*,s	Fruity Pebbles	Oat Squares		
Cocoa Puffs	Corn Flakes*	Grape Nuts	Quaker Oat Bran*,s		
Golden Grahams*	Corn Pops*	Honey Bunches of Oats*,e			
Honey Nut Cheerios*	Cracklin' Oat Bran*	Honeycombs			
Kix	Crispix	Post Fruit & Fiber			
Lucky Charms*	Froot Loops*	Post Raisin Bran*			
Oatmeal Crisp*,s	Frosted Flakes*	Super Golden Crisp			
Oatmeal Raisin Crisp	Frosted Mini Wheats*				
Raisin Nut Bran*	Heartwise*,s				
Total*	Honey Smacks				
Trix*	Kenmei Rice Bran*,s				
Wheaties*	Mueslix				
	Nut & Honey Crunch				
	Nutri Grain*				
	Oatbake*,s				
	Product 19				
	Raisin Bran*				
	Rice Krispies*				
	S. W. Graham*,e,s				
	Special K*				

* Also included in second (ML) estimation step

e New product entry, survived until today

s New product entry, was eventually scrapped

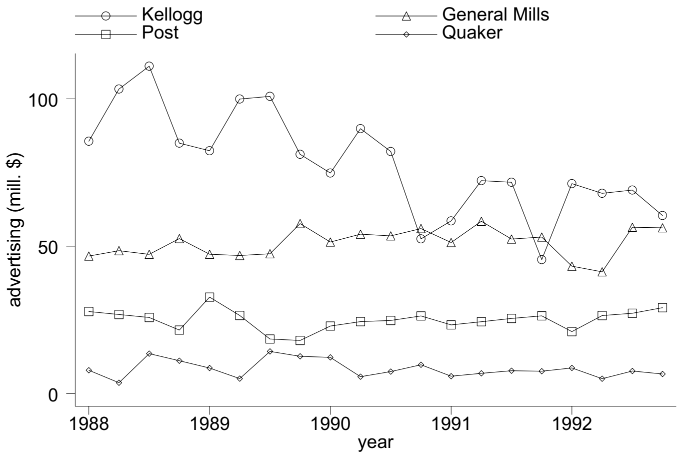

Figure 2: Firm level advertising

and product attributes used in the random coefficients logit model. Note that we truncated the household income and age variables (above \$100,000 and 60 years). This is simply an a priori functional form choice to avoid particularly large or small random coefficients, in particular on the price variable. Such 'outliers' could distort the price elasticity estimates, and lead to numerical instability problems when we re-calculate the price equilibrium.Table 3 shows the first stage regression results. Each regressor is the average price corresponding to product i in market m, where the average is taken over the price of i in all markets $m' \neq m$. This approach yields 20 instruments, each corresponding to one quarter between 1988 and 1992. Table 4 presents the estimation results for several logit models. Comparing models (i) and (ii), we see that the price coefficient almost doubles in absolute size if an IV estimation approach is used over the OLS estimator. The price coefficient is even larger in model (iii), where only the 25 largest breakfast cereals are used instead of the full sample of 56 cereals. This sample of 25 cereals is identical to the sample used by Nevo (2001). On the other hand, as discussed in the estimation section, the estimated median price/cost

Table 2: Sample statistics: Product attributes and demographics

	Mean	Median	Std. dev.	Min	Max	No. obs.
Product characteristics[a]						
Sugar, g	26.78	25.70	13.78	0.85	54.50	56
Fiber, g	6.95	5.50	6.49	0.10	35.30	56
Household demographics[b]						
Household income, \$	35,461	30,212	25,588	1,000	100,000	345,307
Age, years	32.2	32	19.0	0	60	345,307

[a] Amount per 100g cereal.

[b] In order to avoid 'outliers' among the random coefficients, the original household income levels from the CPS below \$1,000 and above \$100,000 were truncated; also age levels above 60 years were truncated.

Table 3: First stage results[a]

Average price, 88-1	-1.283	Average price, 90-3	-4.162
	(0.342)		(0.405)
Average price, 88-2	-2.045	Average price, 90-4	-0.705
	(0.395)		(0.208)
Average price, 88-3	-0.814	Average price, 91-1	-3.108
	(0.364)		(0.417)
Average price, 88-4	-0.477	Average price, 91-2	-3.357
	(0.349)		(0.376)
Average price, 89-1	-1.546	Average price, 91-3	-3.664
	(0.296)		(0.409)
Average price, 89-2	-3.126	Average price, 91-4	-1.824
	(0.354)		(0.274)
Average price, 89-3	-0.685	Average price, 92-1	-1.523
	(0.156)		(0.196)
Average price, 89-4	-3.893	Average price, 92-2	-0.318
	(0.348)		(0.090)
Average price, 90-1	-3.948	Average price, 92-3	-1.134
	(0.352)		(0.194)
Average price, 90-2	-2.885	Average price, 92-4	-2.838
	(0.384)		(0.238)
No. obs.	45,991		
R^2	0.941		

[a]The dependent variable is the product price. Product, market, and time specific effects are included. Standard errors are in parentheses.

Table 4: Logit estimates[a]

	(i)[b]	(ii)[c]	(iii)[c,d]	(iv)[c]	(v)[c]	(vi)[c]	(vii)[c,d]
Price	-0.526	-1.052	-1.426	-0.770	-1.053	-1.024	-1.344
	(0.013)	(0.023)	(0.025)	(0.009)	(0.022)	(0.021)	(0.025)
Advertising					0.079	0.043	0.016
					(0.001)	(0.002)	(0.002)
Advertising_{-1}						0.024	0.004
						(0.002)	(0.002)
Advertising_{-2}						0.022	0.006
						(0.002)	(0.002)
Advertising_{-3}						0.030	0.006
						(0.001)	(0.002)
Advertising_{-4}						0.009	0.011
						(0.001)	(0.002)
Sugar				-0.006			
				(0.0003)			
Fiber				0.022			
				(0.001)			
Calories				0.012			
				(0.0003)			
Fat				-0.085			
				(0.002)			
No. obs.	45,991	45,991	22,124	45,991	45,991	34,608	16,804

[a] All regressions include market, time, and (with the exception of (iv)) product specific fixed effects. Standard errors are in parentheses.

[b] OLS estimates

[c] IV estimates

[d] Sample includes only the 25 products used in Nevo (2001).

margin from the random coefficients logit model is virtually identical to the one reported by Nevo. These findings strongly support the claim that a random coefficients model is needed to accurately estimate product specific price elasticities. The full sample contains many health cereals, which, as we have shown, appeal to a different customer segment than the staples or all-family cereals in the smaller sample of 25 cereals. Hence, the average price elasticity is different between the two samples due to the difference in the composition of products. This difference can only be accounted for by the random coefficients model, which relates price elasticities to product attributes. Columns (vi) and (vii) provide some simple evidence for the importance of the long-run effect of advertising. As can be seen, both current advertising and four lagged levels of advertising have a statistically significant impact on demand. The estimated effect sizes are larger from the full sample than from the smaller sample of 25 cereals. This is as expected, because the full sample contains all new cereals, and hence much more variation in the advertising data than the smaller sample.

Appendix E: Additional Simulation Results

This section presents additional simulation results, corresponding to the parameter estimates in table 5 (ii) in the main paper. Figure 3 shows the profit loss function, the fraction of products launched, and the exit rate under different levels of demand uncertainty. Compared to the results in figure 5 in the main paper, we find that the loss in profits due to uncertainty is generally lower. This is as expected, because the (calibrated) per period fixed cost in table 5 (ii) is lower than the (estimated) fixed cost in table 5 (i), and hence also the loss from launching a product that turns out to be unprofitable is lower. Comparing the launch and exit rates, we find only very small differences between the two different models.

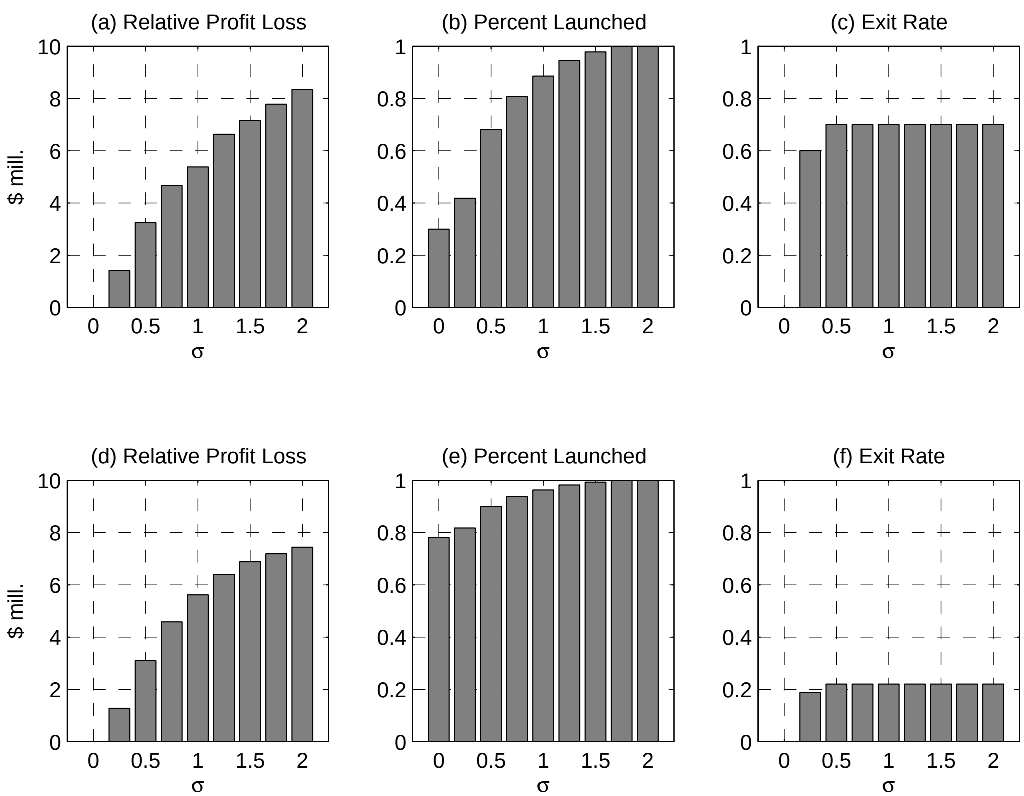

Figure 3: Top row: Products are drawn from the empirical distribution of all newly launched products. Bottom row: Products are drawn from the empirical distribution of all products in the sample.

References

Benítez-Silva, H., G. Hall, G. J. Hitsch, G. Pauletto, and J. Rust (2001): "A Comparison of Discrete and Parametric Approximation Methods for Continuous-State Dynamic Programming Problems," manuscript, Yale University.

Bertsekas, D. P. (1995): *Dynamic Programming and Optimal Control*, vol. 2. Athena Scientific.

Judd, K. L. (1998): *Numerical Methods in Economics*. MIT Press.

Nevo, A. (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69 (2), 307–342.

Nevo, A., and C. Wolfram (2002): "Why Do Manufacturers Issue Coupons? An Empirical Analysis of Coupons and Prices for Breakfast Cereals," *RAND Journal of Economics*, 33(2), 319–339.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (2002): *Numerical Recipes in C++*. Cambridge University Press, second edn.

Rust, J. (1996): "Numerical Dynamic Programming in Economics," in *Handbook of Computational Economics, Vol. I*, ed. by H. M. Amman, D. A. Kendrick, and J. Rust. North-Holland.

Wooldridge, J. M. (2001): *Econometric Analysis of Cross Section and Panel Data*. MIT Press.