Economics and econometrics in modern marketing mix modelling

P.M Cain

Abstract

This article outlines the consumer behavioural foundations of the Marketing Mix Model, ranging from microeconomic competitive structures and on and offline paths to purchase, through to long-term media effects and heterogeneity of consumer response. Such features lead to a high degree of flexibility and realism in marketing mix analysis, helping to maintain its relevance in the modern digital economy.

Several key conclusions for industry practitioners and end-users flow from the discussion. Firstly, if the client-focus is on category level effects of brand specific marketing, standard single equation analysis should be treated with caution. Secondly, offline and online behaviour are inherently linked in the modern digital era. Consequently, focusing solely on direct offline marketing effects on in-store purchase is misleading. Thirdly, conventional model structures are inadequate for quantification of the long-term impact of marketing strategy. Finally, the ability to capture demand response at granular levels is critical in the growing Big Data environment.

Key Words: Attraction model, marketing ROI, long-term effects, Digital media, Path Models, Networks, Vector Autoregression, Unobserved component Models, Hierarchical Bayes.
1) Introduction

The marketing mix model is a popular business tool designed to quantify marketing Return on Investment (ROI), guide the optimal allocation of marketing resources and inform sales forecasting. Understanding the drivers of product demand is central to this process, where the principles of microeconomic consumer theory lie at the heart of the model structure.

Standard demand theory is based on the notion that consumers maximise utility, choosing between competing products based on prices and an income budget constraint. In a market place characterised by differentiated products, marketing strategy plays a key part in this process. As such, traditional marketing mix models are essentially price demand curves augmented to include the role of media in the purchase decision. However, this can only take us so far. In order to capture the full role of marketing in the product sales process, the mix model needs to encompass three additional aspects of consumer behaviour.

Firstly, in-store purchase demand is only part of the story. With the advent of multi-channel marketing in the modern digital era, models need to reflect the full consumer path to purchase. The typical focus is on the role of paid and owned media, with traditional marketing investments stimulating a cycle that starts with natural online search, continuing through to website research and finally onto online and offline sales.

Secondly, in order to successfully explain both call-to-action and brand-building aspects of marketing strategy, it is critical that models are set up to differentiate between short and long-term demand. To address this, modern approaches exploit flexible time series specifications such as the Unobserved Component Model (UCM), directly separating the short and long-run features of the data.

Finally, models must be capable of handling heterogeneity in consumer response. The Big Data revolution has led to a growing wealth of information on individual purchase behaviour, online activity and social media. As the number of marketing response parameters proliferates, conventional fixed-effect estimators struggle to cope. Many MMM studies simplify by aggregating the data or imposing common response patterns across individual units. However, this invokes restrictive assumptions about underlying consumer behaviour (Deaton and Muellbauer, 1980). Consequently, advanced techniques such as Hierarchical Bayesian estimators are often preferred, allowing differential response across consumers in the face of ever-expanding data sets.

In this paper, we develop these issues in detail. Section 2 sets out the microeconomic foundations of the marketing mix model, ranging from standard single-equation forms through to demand system approaches. Section 3 discusses how the mix model has evolved to embrace digital media. Section 4 looks at how the model structure can encompass long-term brand and marketing effects. Section 5 then examines how the model deals with disaggregated data sets and granular consumer response. Section 6 concludes.
2) Classical microeconomic foundations

Successful Marketing Mix Modelling (MMM) requires a solid understanding of the product demand curve. At its most basic, this is defined as the relationship between quantity demanded and price. In perfectly competitive markets, price conveys all the information consumers need to make their choices. In imperfectly competitive markets, however, product differentiation takes on a key role, driven - and bolstered - by a range of non-price related marketing strategies. Under these circumstances, all marketing mix models are effectively price demand curves augmented to include additional media and economic variables that act to shift the demand curve in price-quantity space.

2.1) Theoretical forms

The conventional approach to MMM analysis focuses on ‘loose’ demand curves for selected items and/or brands in the manufacturer’s portfolio and typically take the following form:

\[ \ln S_{it} = \alpha_i + T_i + \delta_{it} + \gamma_{ij} \ln P_{jt} + \sum_{k=1}^{M-1} \beta_{ijk} \ln X_{kit} + \varepsilon_{it} \]

Which stipulates that sales of product \( i \) over time \( t \) are a multiplicative function of its own price \( P_i \), competitor prices \( P_j \) and a set of own and competitor marketing and economic driver variables \( X_{kit} \). The demand equation is completed with an intercept \( \alpha_i \), trend \( (T_i) \), seasonal index \( (\delta_i) \) and an error term \( \varepsilon_i \). The intercept is equal to the mean of the sales data - net of the parameter weighted means of the explanatory variables - and equivalent to the expected level of non-marketed product sales. This is often referred to as base sales. The trend term caters for any observable ‘drift’ present in the base over time and the seasonal index caters for regular ‘time-of-year’ factors and period-specific holiday effects. The error term represents all unexplained factors.

If the focus is on one product at a time, single equation forms are perfectly adequate. However, they do suffer from two inter-related drawbacks. Firstly, if the model is applied to several competing products in one group or ‘category’, it is perfectly possible that total volume gains are either greater than or less than total volume losses. This is typically interpreted as category growth or shrinkage respectively. However, in reality it is predominantly a consequence of the fact that sets of single equations are unrelated and do not ‘add up’. As such, they are inconsistent with the budget constraint of conventional microeconomic demand theory (Deaton and Muellbauer, 1980), telling us nothing about genuine category effects of product marketing.\(^2\)

---

\(^1\) Additive linear forms are also used. However, the multiplicative model is often chosen due to implied non-linear relationships. For example, demand drops exponentially to zero as price approaches infinity and advertising exhibits diminishing returns as weight increases. These, together with the implied synergies between the variables, are deemed desirable properties.

\(^2\) Violation of adding up indicates that single-equation functional forms give distorted expenditure patterns across the modelled products that are greater or less than total expenditure across the group.
Secondly, we cannot accurately define the source of incremental volume to each element of the marketing mix: specifically, how much is due to substitution from other brands and how much is due to category expansion effects? This distinction is important if manufacturers are seeking empirical evidence to justify increased shelf-space in-store.

To overcome these problems, closer adherence to underlying microeconomic utility theory is required where consumers choose between competing products based on price, marketing and an income (budget) constraint. This structure is more accurately represented using information across the full product consideration set, treating the defined category as a single unit. This leads to simultaneous-equation demand system approaches, delivering internally consistent estimates of volume gains and losses together with meaningful category expansion effects of brand specific marketing. The result is more accurate price elasticities, marketing ROI and budget allocation facilitating the manufacturer-retailer relationship (Cain, 2014a).

Theoretical demand system structures fall into two broad camps. On the one hand, we have the continuous choice models such as Stone’s Linear Expenditure system (Stone, 1954), the Rotterdam model (Theil, 1965, Barten, 1966) and the Almost Ideal Demand System (AIDS) of Deaton and Muellbauer (1980). These systems are based on standard utility maximisation theory and appropriate if we believe consumers choose equilibrium quantities of all products in the choice set (Hausman et al., 1994). On the other hand, there are the discrete choice forms such as the attraction model illustrated in Nakanishi and Cooper (1974) and Fok et al. (2001). These models are rooted in alternative characteristics-based demand theories (Lancaster, 1971) and generally preferred in marketing since consumers typically choose only one product from the consideration set.

2.2) Econometric estimation

Whether we opt for single equation or system approaches to the mix model, it is important to recognise that theoretical model structures such as equation (1) depict contemporaneous relationships between sales, price and marketing variables. This is as far as economic theory can take us in deriving a suitable model form and implicitly assumes that consumers adjust immediately to changes in the driver variable(s). However, factors such as brand loyalty and habit formation suggest sluggish adjustment to points of new equilibrium. In order to resolve this, more flexible dynamic MMM forms are required for estimation purposes.

A general dynamic specification rewrites the static model form as an Autoregressive Distributed Lag (ADL) function in terms of lagged sales together with current and past driver values. In the case of the single equation model (1), we have:

---

3 Specifying marketing effects in relative terms can help here. For example, estimating price in absolute and relative terms provides an elasticity split into two components: a relative demand response when price of good $i$ changes relative to all other goods in the group and a matched demand response when prices of all products in the group move together. Although analogous to a separation of substitution and category level effects, this approach is still inconsistent with the adding up constraint.
\[
\ln S_{it} = \alpha_i + T_i + \delta_i + \sum_{l=1}^{T} \gamma_{lj} \ln P_{jit} + \sum_{l=1}^{T} \sum_{k=1}^{M-1} \beta_{ikj} \ln X_{kit} - l + \sum_{l=1}^{T} \lambda_i \ln S_{it-l} + \epsilon_{it}
\]

Equation (2) depicts a dynamic relationship and is intended to capture the short to medium term effects of price and marketing investments as consumers adjust to new levels of the relevant driver(s). Analogous specifications can also be applied to demand system approaches. However, the final lag length \(l\) of the price, marketing and sales variables is purely data driven: a priori economic theory does not tell us precisely how consumers adjust. Nevertheless, many types of dynamic adjustment patterns are embedded within this general functional form.

One basic type of adjustment path is a restricted form of the distributed lag model – nested in equation (2) – with a constrained lag \(l\) structure applied solely to the advertising variable(s). This is equivalent to the Adstock concept, introduced into the marketing literature by Broadbent (1979) and represents the conventional approach to incorporating dynamics into the marketing mix model. The idea is intended to capture the direct current and future effects of advertising, where a portion of the full effect is felt beyond the period of execution due to media retention and the product purchase cycle.

Alternative dynamic structures are based on the partial adjustment model (Chow, 1957) and focus solely on current period \(X\) variables and lagged sales in (2). Here, consumers are assumed to partially adjust - at a rate determined by \(\lambda\) - towards a desired or equilibrium sales level following a change in the marketing variables. Others are based on the error correction form (Davidson et al., 1978), where both current and lagged \(X\) variables appear together with lagged sales. Here, the hypothesis is that an equilibrium relationship exists between sales and the levels of chosen drivers. Deviations from equilibrium are then embedded - and thus corrected for - in the short-run dynamic adjustment process.

3) Consumer off and online path to purchase

With the advent of digital marketing, final product purchase is typically the end result of an off and online decision making process (Cain, 2014b), where classical microeconomic utility theory and product pricing is only one part of the story. Consequently, research-based consumer journey theories have arisen that build on the demand models of Section 2 to provide a fuller explanation of the final purchase decision.

3.1) Theoretical structure

A common hypothesis contends that traditional marketing investments stimulate a journey that starts with natural online search, continues through to website research and finally

---

4 Standard Adstock analysis assumes that the distributed lag coefficients decline geometrically – with maximum impact in the current period. Alternative lag structures may be imposed, such as Polynomial Distributed lags (Almon, 1965), where maximum response to advertising can occur after the period of deployment mimicking advertising ‘wear-in’ effects.
onto online and offline product purchase. This structure is intended to capture how holistic paid media strategies drive owned media and the brand content experience. Social or earned media then deals with how consumers talk about and share the content experience, reflected in additional steps such as Twitter feeds and participation in Facebook discussions or fan pages. This sequence is illustrated in Figure 1, where digital and traditional marketing investments impact each step of the journey with controls for all external and macroeconomic influences.5

**Figure 1: Offline-online consumer journey**

![Figure 1](image)

### 3.2) Econometric estimation

The consumer journey approach to MMM essentially constitutes a network of equations ‘nested’ within the final product demand curve structures of Section 2. In econometric terms, this constitutes a structural equation model, where dependent (explained) variables also serve as independent (explanatory) drivers. This poses a problem for standard regression methods, where valid causal inference requires that each explanatory variable is exogenous or fixed outside the model. Dependent variables are determined inside the system and defined as endogenous. If used as contemporaneous explanatory variables, this leads to an identification problem, confounding correlation and causation. The result is biased estimation of the system parameters. Various econometric techniques are deployed to handle this, where the causal relationships are modelled subject to a chosen identification scheme. Three distinct approaches are typically used.

#### 3.2.1) Path models

Path models are basic networks comprising sets of regression equations with an assumed causal ordering across outcomes. This gives a recursive ‘triangular’ structure, equivalent to a set of identifying restrictions.6 The sequential process illustrated in Figure (1) is an example

---

5 Figure 1 is a form of Directed Acyclical Graph (DAG). Cases where final sales subsequently drive natural search constitute a circular system where no clear causal path can be defined and typically excluded.

6 Note that identification requires that the error terms of each equation must also be independent of all others in all time periods.
of a simple off and online purchase journey represented by the following set of conditional probability statements.

\[
\begin{align*}
P(\text{Natural Search}_t | x_t) \\
P(\text{Webtraffic}_t | \text{Natural Search}_t, x_t) \\
P(\text{Facebook}_t | \text{Webtraffic}_t, \text{Natural Search}_t, x_t) \\
P(\text{Twitter}_t | \text{Facebook}_t, \text{Webtraffic}_t, \text{Natural Search}_t, x_t) \\
P(\text{Sales}_t | \text{Twitter}_t, \text{Facebook}_t, \text{Webtraffic}_t, \text{Natural Search}_t, x_t)
\end{align*}
\]

Model (3) is estimated as a system of equations where \( x_t \) denotes a set of exogenous price, marketing and macroeconomic explanatory variables. Starting with natural online search, the journey follows a path where each step is pre-determined and handed down as a driver into the next level, culminating in the final off and/or online sales demand equation.\(^7\)

### 3.2.2) Bayesian Networks

Bayesian networks represent sets of conditional dependencies between groups of random (outcome) variables. This generalises the path approach by looking at the combination of all possible chains that can explain the final outcome. For example, the process illustrated in Figure 1 is just one possible route to off and online sales with a given overall likelihood. However, there are many others such as Twitter driving Facebook, Facebook directly driving Web Traffic, or social media indirectly impacting sales via Web Traffic. A set of key causal structures can be found using machine learning techniques or score based approaches. Plausible candidates are chosen based on economic and relative likelihood criteria, with network parameters estimated by either frequentist or Bayesian methods.\(^8\) The final model for sales can be viewed as a weighted combination of each estimated network.

### 3.2.3) Structural Vector Autoregression (SVAR)

Vector Autoregression (VAR) models (inter alia Lutkepohl, 1991 and Juselius, 2006), specify Figure 1 as a system of regression equations, where each endogenous outcome variable \( z_t \) is written in terms of the past behaviour of all outcome variables and current and lagged exogenous variables \( x_t \). This constitutes a vector analogue of equation (2), written as:

\[
z_t = A_1 z_{t-1} + A_2 z_{t-2} + \ldots A_l z_{t-l} + \Phi_1 x_t + \ldots \Phi_l x_{t-l} + \varepsilon_t
\]

Where \( A_1, \ldots, A_l \) and \( \Phi_1, \ldots, \Phi_l \) denote matrices of regression parameters for all endogenous (outcome) and exogenous (marketing) driver variables respectively and \( \varepsilon_t \) is a covariance matrix of error terms.

\(^7\)Typically, the set of exogenous \( x_t \) variables differs across equations and endogenous regressors can be removed from some steps due to collinearity or insignificance. Note too that causal effects can be direct or indirect where the effect is mediated by another outcome variable.

\(^8\) Bayesian Networks are so called owing to the use of Bayes’ rule when conducting probabilistic inference, rather than the statistical approach to parameter estimation.
The VAR approach is widely documented in the economics and econometrics literature, with many applications in macroeconomics (inter alia Hamilton, 1994 and Hendry et al. 1991) and marketing (inter alia Dekimpe and Hanssens, 1995, Trusov et al. 2009). However, as they stand, VAR models such as (4) are reduced form equations, where the contemporaneous (causal) relationships between the outcomes are not modelled in any way. All such relationships are absorbed into the error covariance structure $\epsilon_t$, leading to an identification problem manifest in significant off-diagonal error matrix correlations.

Structural Vector Autoregression (SVAR) resolves the problem by removing cross-correlation in the system error matrix $\epsilon_t$. Many examples can be found in the macroeconomics literature (inter alia, Blanchard and Quah, 1991, Crowder et al., 1999 and Pagan and Pesaran, 2008), where various identification approaches are applied. The simplest option imposes a Cholesky decomposition, enforcing zero restrictions on all off-diagonal matrix elements. This is essentially analogous to the path model, giving a recursive structure and an assumed causal ordering between the variables. Alternatively, order-agnostic solutions can be provided through generalised impulse response analysis (Pesaran and Shin, 1998). Other options include model restrictions based on economic theory or explicitly modelling the causal relationships using instrumental variable techniques (Juselius, 2006).

4) Long-term consumer behaviour

A common charge levied at all types of conventional MMM analysis is the inability to accommodate long-term advertising effects, thereby underestimating true media effectiveness. This is unsurprising given standard demand structures such as equation (2). Here, product demand is essentially assumed to fluctuate around a constant mean level $\alpha$, with no allowance for persistent and lasting evolution in brand tastes. A deterministic drift factor $T$ is often added to proxy long-term behaviour, but is wholly deterministic and user-imposed. In MMM terms, this implies fixed or deterministic baselines by construction, telling us nothing about long-term brand evolution. In reality, however, the baseline is essentially a dynamic construct impacted by many factors including media investments.

4.1) Theoretical structure

Base sales evolution indicates the extent to which new purchasers are converted into loyal consumers – through persistent repeat purchase behaviour and lasting shifts in consumer product tastes. This, in turn, can lead to shifts in price elasticity as stronger equity reduces demand sensitivity to price change. Brand perceptions can play a key role in this process, driving product tastes and repeat purchase behaviour. Marketing strategy, in turn, works directly on brand awareness, consideration and perceptions. This reasoning creates the flow

---

9 VAR models are often used to test for Granger causality between variables (Granger, 1988). However, this is a specific form of predictive causality based solely on past values of the variables concerned. It does nothing to address the identification problem at the heart of causal analysis in structural systems.

10 Note that brand perceptions are also forged by product experience implying a reverse causality, where base sales and perceptions can drive each other.
illustrated in Figure 2, where media investments are linked to variation in base sales and price elasticity via consumer brand tracking metrics.

**Figure 2: Long-run advertising eco-system**

4.2) Econometric estimation

Estimation of the long-term advertising eco-system in Figure 2 requires a new two-stage approach to the mix model. Firstly, we need to extend the frameworks outlined in Sections 2 and 3 to encompass baseline evolution as part of the basic model structure. Secondly, we require an auxiliary model process that can link extracted baselines to primary research consumer tracking data. The next two sections outline these steps in more detail.

4.2.1) Estimating baseline evolution

In order to measure both marketing incremental and baseline evolution, modern approaches to the mix model re-cast the conventional structures in a more flexible Unobserved Component Model (UCM) form (Cain, 2008). Using the single equation model (1) as an example, the general specification is as follows:

\[
\begin{align*}
\text{Ln} S_{it} &= \mu_{it} + \delta_{it} + \sum_{l=1}^{T} \gamma_{ij} \ln P_{ijt} + \sum_{k=1}^{M-1} \beta_{ijk} \ln X_{kit} + \epsilon_{it} \\
\mu_{it} &= \mu_{it-1} + \lambda_{it-1} + \eta_{it} \\
\lambda_{it} &= \lambda_{it-1} + \xi_{it} \\
\gamma_{it} &= \gamma_{it-1} + \nu_{it} \\
\delta_{it} &= -\sum_{j=1}^{p-1} \delta_{it-j} + \kappa_{it}
\end{align*}
\]

1(a), 5(a), 5(b), 5(c), 5(d)
Equation 1(a) replaces the intercept $\alpha_i$ in equation (1) with a time varying (stochastic) trend $\mu_t$, comprising two components described by equations 5(a) and (b). This is known as the local linear trend model (Harvey, 1989). Equation 5(a) allows the underlying sales level to follow a random walk with a growth factor $\lambda_i$, analogous to the conventional trend term $T_i$. Equation 5(b) allows $\lambda_i$ itself to follow a random walk. The variables $\eta_i$ and $\xi_t$ represent two mutually uncorrelated normally distributed white-noise error vectors with zero means and covariance matrices. Depending on the estimated values of the covariance variances, the system can accommodate both stationary and non-stationary product demand allowing the data to decide between them.

Equation 5(c) allows for evolution in price sensitivity for product $i$ ($\Upsilon_i$) over the sample with a random error $\nu_\iota$ with mean 0 and covariance matrix $\sigma^2_\iota$. Equation 5(d) specifies seasonal effects, which are constrained to sum to zero over any one year. Stochastic seasonality is allowed for using dummy variables, where $p$ denotes the number of seasons per year, $\delta_t$ is the seasonal factor corresponding to time $t$ and $\kappa_i$ is a random error with mean 0 and covariance matrix $\sigma^2_\kappa$. If the latter is zero, then seasonality is deterministic.

Examples of the dynamic UCM formulation can be found in Hunt et al. (1999), Moosa and Baxter (2002) and Cain (2005). Applied to marketing mix analysis, it provides a fully flexible modelling framework. In the first place, product demand is directly decomposed into long-term and short-term components. Specifically, $\mu_t$ in equation 5(b) measures long-run changes in demand arising through shifts in underlying consumer tastes, leaving the $\beta_{ijk}$ parameters of equation 1(a) to accurately measure short-run demand changes due to marketing activity. The result is a more realistic split into base and incremental volumes and more accurate short-term ROI calculation.

Secondly, the framework can accommodate the conventional short-term dynamics typically incorporated in the marketing mix model. For example, advertising distributed lag effects can be incorporated in equation 1(a) in the form of conventional Adstock variables. Improved short and medium-term dynamic specification here provides a cleaner read on the long-term evolving component of the sales series.11

4.2.2) Modelling baseline evolution

Having isolated the short-term impacts, the extracted trend component allows us to focus specifically on base-shifting (long-term) effects.12 The structure of Figure 2 represents a network model, similar to Figure 1, comprising sets of simultaneous outcomes and exogenous marketing and control variables. The key difference here, however, is that for meaningful long-term effects to exist the variables involved must be evolving (non-stationary). Under these circumstances, the natural choice of estimation technique is the

---

11 Specifying the underlying demand level to evolve as a (non-stationary) random walk essentially caters for situations where the explanatory variables cannot fully explain the level of sales. Since the random walk form contains elements at all frequencies, it can also reflect missing short and medium term information.

12 Note that an alternative approach to evaluating long-term effects incorporates media investments directly into the trend evolution equations 5(a) and (b) (Osinga et al. 2009).
Cointegrating Vector Auto Regression (VAR) model (Johansen, 1996, Juselius, 2006). For example, consider the following VAR specification of Figure 2:

\[ z_t = A_1 z_{t-1} + \cdots + A_k z_{t-l} + \Phi_1 x_t + \cdots + \Phi_k x_{t-l} + \varepsilon_t \]  

(6)

Where \( z_t \) is a vector of time series data on base sales, consumer tracking metrics, price elasticity and control variables and \( x_t \) is a matrix of current and lagged exogenous (stationary) marketing explanatory variables. Equation (6) is then re-written in Vector Error Correction (VECM) form:

\[ \Delta z_t = \Gamma_1 \Delta z_{t-1} + \cdots + \Gamma_{k-1} \Delta z_{t-k+1} + \Pi z_{t-1} + \Phi_1 x_t + \cdots + \Phi_k x_{t-k} + \varepsilon_t \]  

(7)

Equation (7) depicts a short-term (first differenced) equation for each endogenous variable \( z_t \), in terms of lagged differences, lagged levels and marketing drivers \( x_t \). In its current form equation (7) is unbalanced, comprising stationary (differenced) and non-stationary (level) data. This can be resolved if there are stationary linear combinations of \( z_t \), where the long-run \( \Pi \) matrix is of reduced rank and the variables cointegrate. Under these circumstances, equilibrium relationships exist: that is, relationships which are restored (error-correct) when disturbed such that the outcome series follow long-run paths together over time. Equation (7) is then written:

\[ \Delta z_t = \Gamma_1 \Delta z_{t-1} + \cdots + \Gamma_{k-1} \Delta z_{t-k+1} + \alpha_1 \beta^t z_{t-1} + \Phi_1 x_t + \cdots + \Phi_k x_{t-k} + \varepsilon_t \]  

(8)

Where the \( \Pi \) matrix is factorised into \( r \) cointegrating relationships \( \beta \) and associated error-correction parameters \( \alpha \). In the context of Figure 2, we could expect three cointegrating relationships for base sales, brand perceptions and price elasticity with two common trends driving the non-stationary properties of the system.

As it stands, all the variables in \( z_t \) are allowed to be endogenous. However, this is unlikely for certain variables such as macroeconomic control factors. Consequently, it is often preferable to partition \( z_t \) into genuine endogenous variables \( y_t \) and weakly exogenous variables \( w_t \).\(^{13}\) We may then write:

\[ \Delta y_t = \Gamma_0 \Delta w_t + \Gamma_1 \Delta z_{t-1} + \Gamma_{k-1} \Delta z_{t-k+1} + \alpha_1 \beta^t z_{t-1} + \Phi_1 x_t + \cdots + \Phi_k x_{t-k} + \varepsilon_t \]  

(9)

Equation (9) then allows us to test the extent to which the marketing factors load into each of the first differenced equations – and cumulate permanently into the levels of the endogenous variables.\(^{14}\) We are testing for two key long-term media effects here. Firstly, the impact on the consumer tracking data and the extent to which this indirectly cumulates permanently into the level of base sales via the cointegrating relationships. Secondly, the degree to which media directly drives price elasticity over time.

\(^{13}\) Note that the weakly exogenous variables can still play a role in determining the long-run cointegrating relationships.

\(^{14}\) This is analogous to impulse response analysis (inter alia, Sims, 1980). However, standard impulse response functions simulate the dynamic system-wide impact of shocks to the error structures of the relevant variables. Clearly this requires we treat the media variables as endogenous in Equation (9) (Cain, 2010). This can be problematic for GRP advertising data for example, which tends to follow discrete bursts over time.
5) Consumer heterogeneity

A central feature of demand behaviour is the typically high degree of heterogeneity that exists across individual consumers. For more informative and actionable MMM analysis, it is important that such differences are reflected in the modelling process. In practice, this essentially amounts to defining the model structures of sections 2, 3 and 4 over increasing levels of data disaggregation. In this final section, we look at common disaggregated data structures used in marketing analytics, together with econometric techniques capable of capturing heterogeneous consumer response.

5.1) Data structure

Data sets in marketing analytics are typically defined over products (depth), time (frequency) and the number of variables involved (variety). Figure 3 illustrates a typical case, where market level sales are split into cross-sections with each block depicting a time series of consumer demand for each product-channel-regional combination. Sales channels may be further divided into stores or individual customers. Coupled with information on pricing, promotion, offline media, customer attributes and profiles, data sets start to expand significantly.

**Figure 3: Longitudinal data structure**

The digital revolution has only exacerbated this problem, leading to a proliferation in the volume and variety of online information such as web traffic by source, natural and paid search behaviour by platform, display exposure and social media feeds – much of which is available at individual consumer level by day. Merged with offline data to create a holistic view of consumer demand creation, this is the essence of Big Data in marketing analytics.
5.2) Econometric estimation

In order to handle increased data size and complexity, MMM analysis often aggregates the data to more manageable proportions. However, this loses the value that modern data granularity has to offer. In the next three sections, we look at some common approaches that help resolve this *dimensionality* problem. Each approach constitutes alternative methods of data pooling, enabling a significant reduction in the number of specified parameters.  

5.1) Conventional disaggregated data analysis

Analysis of disaggregated data sets is very common in marketing analytics. Rather than estimate one product, one cross section at a time, it is preferable to take advantage of the longitudinal structure, stack the data and estimate all dimensions simultaneously. Typically, we are seeking to quantify cross sectional deviations (interaction effects) from the market mean or base level of response (main effect). This can be seen as analogous to data mining techniques such as CHAID and helps improve targeting of marketing investments. To illustrate the issues involved, consider the data structure for Figure 3 expressed in stacked form as follows:

\[ \begin{align*}
\mathbf{y}_{it} &= \mathbf{x}_{it} \mathbf{\beta}_i + \mathbf{\varepsilon}_{it} \\
\end{align*} \tag{10} \]

Where \( i = 1-N \) denotes the defined cross-sectional unit, \( t = 1-T \) denotes the time period, \( \mathbf{y}_{it} \) denotes a vector of dependent variables for product or brand sales, \( \mathbf{x}_{it} \) denotes a row vector of \( K \) current and lagged explanatory variables for cross-section \( i \), \( \mathbf{\beta}_i \) is a \( K \)-vector of response coefficients and \( \mathbf{\varepsilon}_{it} \) represents a vector of error terms. Specific intercepts or fixed effects are typically added to each row to account for mean cross-sectional differences.  

Appropriate estimation of (10) depends on the properties of the error structure, both within and across cross-sections. Classical Ordinary Least Squares (OLS) requires that the error covariance matrix of the \( i^\text{th} \) cross section satisfies the standard assumptions of constant variance and zero serial correlation:

\[ E(\mathbf{\varepsilon}_{it} \mathbf{\varepsilon}_{it}') = \delta_i^2 I_T = \Pi_i \tag{11} \]

With zero contemporaneous error correlation across cross-sections:

---

15 Methods discussed constitute common econometric approaches to dealing with consumer heterogeneity. This is distinct from other simulation based approaches such as Agent Based Modelling (Rand and Rust, 2011), which seeks to explain complex macro behaviour via bottom-up rule-based heterogeneous agent behaviour.

16 Note that equation (10) is the panel data analogue of equation (2) where the fixed effects are equivalent to the intercept term. Allowing these terms to evolve gives the panel data version of model 1(a)-5(b).
Ordinary Least Squares (OLS) estimation of (10) provides the coefficient vector \( \beta_i \), with marketing response estimates specific to each cross-section \( i \). In circumstances where (11) and (12) do not hold, such as in the presence of heteroscedastic and/or autocorrelated errors, Generalised Least Squares (GLS) approaches are typically applied. This uses OLS to estimate the relevant error structures and transform the data such that (11) and (12) are then applicable.\(^{17}\)

**5.2) Big Data methods**

As modern data sets burgeon in size, obtaining reliable detailed marketing response information can be challenging. Full interaction time series-cross-sectional approaches such as model (10) are often unstable as the number of dimensions and parameters increase, delivering many zero and/or incorrectly signed effects. A natural solution to this dimensionality problem is to reduce the number of estimated parameters, thereby increasing the available degrees of freedom.

The simplest approach is to aggregate and reduce the size of the data set. This is a fairly common practice: Big Data challenges often revolve around storage issues in the first instance but, once processed, data are aggregated to simplify analysis. In marketing analytics, behaviour is often summed over time to a weekly frequency and attention is focused on average relationships across brands and sales channels.

Valid aggregation, however, requires quite specific assumptions about consumer behaviour (Deaton and Muellbauer, 1980). Furthermore, if the demand relationships are non-linear at granular levels then application of these same forms to linearly aggregated data results in bias (Christen et al., 1997). Consequently, alternative methods are required to handle increasing data size and granularity that obviate the need for aggregation over product and consumer dimensions.\(^{18}\)

**5.2.1) Classical pooling**

The most basic approach is to pool the data. This provides a single average response coefficient \( \beta \) for each relevant explanatory variable in model (10). The downside is that this ignores response heterogeneity over cross-sections and is of little use to media planners and budget holders seeking guidance on media targeting at regional level for example. This can be remedied by regional pooling, but at the cost of increasing the number of parameters whilst still imposing homogeneity across products and consumers.

\(^{17}\) Incorporating digital media into equation (10) requires a model of the off and online consumer purchase journey and appropriate estimation techniques as set out in Section 3.

\(^{18}\) Given the increased storage needs and computational complexity involved in Big Data sets, these methods often deploy *sparse matrix* techniques to facilitate large scale modelling and optimisation.
5.2.2) Hierarchical Bayes

A more flexible technique is to introduce random coefficients in the form of a Hierarchical Bayesian structure. Equation (10) is the typical structure of the classical or frequentist approach to statistics, where the model parameters are regarded as unknown fixed quantities to be estimated from the data. In the Bayesian approach, parameters are viewed as unknown outcomes of a random process determined by another higher level joint distribution. In the context of Figure 3, this assumes that each cross sectional coefficient is drawn from a population distribution shared by all the cross-sections. This is a strong assumption, but leads to a dramatic reduction in the number of estimated parameters. For example, consider model (5) re-written as follows:

\[ y_{it} = x_{it}\beta_i + \varepsilon_{it} \]  \hspace{1cm} (10a)
\[ \beta_i = \beta + v_i \]  \hspace{1cm} (13)
\[ y_{it} = x_{it}\beta + u_{it} \]  \hspace{1cm} (14)
\[ u_{it} = x_{it}v_i + \varepsilon_{it} \]  \hspace{1cm} (15)

Where equation 10(a) represents each cross-sectional (micro) model and equation (13) represents a higher level (macro) distribution for coefficients \( \beta_i \) with mean \( \beta \) and error \( v_i \) - denoting the random spread around the mean. It is this micro and macro view that gives the model its hierarchical structure. Combining 10(a) and (13) leads to model (14), with a composite error term \( u_{it} \). Covariance matrix (11) then becomes:

\[ E(u_{it}u_i') = E[(x_{it}v_i + \varepsilon_{it})(x_{it}v_i + \varepsilon_{it})'] = \sigma_i^2 I_T + x_i\Gamma x_i' = \tilde{\Gamma}_i \]  \hspace{1cm} (11a)

The error covariance matrix of each cross-section is now a function of both the variance (\( \sigma_i^2 \)) and the parameter spread over cross-sections (\( \Gamma \)). The balance between the two determines the estimated \( \beta_i \) coefficients and reveals the distinctly Bayesian nature of the approach. A high value of \( \sigma_i^2 \) relative to \( \Gamma \) implies relatively imprecise cross-sectional estimates. Consequently, the data have little to offer and the individual \( \beta_i \) values are shrunk towards the pooled (prior) mean value. Conversely, where \( \sigma_i^2 \) is low relative to \( \Gamma \) so the sample data are more informative and the cross-section specific estimates dominate with minimal shrinkage. In this way, the Hierarchical Bayesian estimator is essentially a weighted average of the pooled and cross-sectional estimates.\(^{19}\)

The hierarchical structure of model 10(a) – (15) indicates that coefficients by cross-section can be obtained simply through knowledge of the mean and variance of the macro

\(^{19}\) Note that as the number of time series observations increases, so the Hierarchical Bayesian result converges to the classical cross-sectional specific estimates.
distribution (13) plus the error variance $\sigma_i^2$ of the micro model 10(a). This is far more parsimonious than the classical approach and often seen as a distinct advantage in the face of modern large data sets. Parameter estimation sets the mean of the macro distribution to the market level (pooled) estimate and the variance is derived from the global spread of the individual cross-section parameter estimates. Alternatively, where systematic regional differences are known to exist, it is preferable to set regional mean priors via pooling across products and chains, with variances estimated using intra-regional coefficient spreads.

5.2.3) Attribute based models

Our third example is based on the economics of how consumers shop for products. Marketing mix models are based on conventional microeconomic demand theory, where consumer preferences are defined over the individual products themselves. However, an alternative approach defines preferences across higher level product attributes and characteristics (Lancaster, 1971). For example, the television category can be divided into brand, screen size and technology and further divided into brand name, dimensions and LCD/LED/Plasma/3D. Provided that there is a sufficient level of commonality in attributes and characteristics across the category, a complete product (SKU) level data set can be fully described over a significantly reduced number of dimensions.

An application of the characteristics approach to demand in the marketing literature can be found in Fader and Hardie (1996) and represents a highly efficient method of parameter reduction. In practice, it is common to combine this approach with the Hierarchical Bayesian technique for even greater parsimony as data sets expand in size.

6) Concluding remarks

This article has outlined the consumer behavioural foundations of MMM analysis, ranging from microeconomic competitive structures and the incorporation of digital and social media through to measurement of long-term effects and heterogeneity of consumer response. All such components lead to increased flexibility and realism, helping to maintain the relevance of MMM in the modern digital economy. Several practical conclusions for industry practitioners and end-users emerge from this.

---

20 The Hierarchical Bayesian model 10(a)-15 assumes that all cross sectional differences are driven by chance. However, practitioners typically interpret and use the predicted coefficients in exactly the same way as systematic fixed coefficients. Strictly speaking this is invalid, but can be partially rectified by randomising around a set of pre-defined clusters or levels rather than at the national level. This amounts to specifying an additional equation (13) for each defined cluster, increasing the number of parameters to estimate.

21 This estimation approach is known as Empirical Bayes, where the full-interaction estimator (10) is used to calculate cross sectional specific coefficients. However, these estimates serve only to calculate the covariance matrix $\Gamma$.

22 Pure Bayesian approaches (Rossi et al., 2005) set priors independently of the data and represent an increasingly popular method of introducing user-control into marketing mix modelling. A prior for $\beta$ allows us to set the mean value of the macro distribution to externally given values. This is particularly useful if we wish to constrain parameters to be positive or negative and/or set values consistent with previous studies. Priors for the coefficient dispersion (covariance) matrix $\Gamma$ then allow control over the degree of shrinkage around the mean.
If the focus is on category level effects of brand specific marketing, single equation MMM analysis should be treated with caution. Correct inference requires a simultaneous equation demand system approach that separates marketing effects into sum-constrained substitution and category level income effects.

MMM analysis that focuses solely on traditional direct marketing impacts on in-store purchase is insufficient in the modern digital era. The advent of multi-channel off and online marketing, has forced modellers to re-think the basic economic structure in order to incorporate digital media such as paid search and online display. This involves a recognition of the off and online consumer journey, requiring more advanced econometric techniques where the causal relationships are modelled subject to chosen identification schemes.

Conventional MMM structures are inadequate for quantification of the long-term impact of marketing strategy. Traditional dynamic analysis focuses solely on constrained distributed lags in the form of Adstock variables, ignoring the impact of media campaigns on long-term brand tastes. Meaningful long-term analysis requires a model framework that simultaneously estimates both short-term incremental and long-term baseline evolution.

Finally, the ability to capture demand response at granular levels is critical for targeted advice on above and below the line marketing strategy. In order to handle the increasing size of modern data sets, MMM analysis often adopts simple pooling techniques or aggregates information to more manageable proportions. However, this loses the value that modern data granularity has to offer. Advanced techniques such as Hierarchical Bayesian analysis represents a common solution to this problem, allowing stable estimation of a proliferating number of detailed marketing response parameters.
References


Cain, P.M. (2008), ‘Limitations of conventional marketing mix modelling’, *Admap, April* pp 48-51


Cain, P.M. (2008), ‘Limitations of conventional marketing mix modelling’, *Admap, April* pp 48-51


