ONLINE APPENDIX A: DERIVATIONS FOR SECTION 2

Without loss of generality, assume $BC_{it} > 0$. Procyclical spending is beneficial if:

$$\text{(A.1)} \qquad \varepsilon_{\text{weight}} = \frac{\partial s_{it}}{\partial BC_{it}} \frac{BC_{it}}{s_{it}} = \frac{\partial\left(\frac{W_{it}}{\sum_{j=1}^{n} W_{jt}}\right)}{\partial BC_{it}} \frac{BC_{it}}{\left(\frac{W_{it}}{\sum_{j=1}^{n} W_{jt}}\right)} > 0.$$

Working through this derivative, we obtain:

$$\text{(A.2)} \qquad \frac{\partial\left(\frac{W_{it}}{\sum_{j=1}^{n} W_{jt}}\right)}{\partial BC_{it}} \frac{BC_{it}}{\left(\frac{W_{it}}{\sum_{j=1}^{n} W_{jt}}\right)} = \frac{\frac{\partial W_{it}}{\partial BC_{it}} \cdot \sum_{j=1}^{n} W_{jt} - W_{it} \cdot \frac{\partial \sum_{j=1}^{n} W_{jt}}{\partial BC_{it}}}{\left(\sum_{j=1}^{n} W_{jt}\right)^2} \frac{BC_{it}}{\left(\frac{W_{it}}{\sum_{j=1}^{n} W_{jt}}\right)} =$$

$$\frac{\frac{\partial W_{it}}{\partial BC_{it}} \cdot \sum_{j=1}^{n} W_{jt} - W_{it} \cdot \frac{\partial W_{it}}{\partial BC_{it}}}{\left(\sum_{j=1}^{n} W_{jt}\right)^2} \frac{BC_{it}}{\left(\frac{W_{it}}{\sum_{j=1}^{n} W_{jt}}\right)} = \left(\frac{\sum_{\substack{i'=1 \\ i' \neq i}}^{n} W_{i'}}{\sum_{j=1}^{n} W_{jt}}\right) \frac{BC_{it}}{W_{it}} \frac{\partial W_{it}}{\partial BC_{it}} = c_{it} \frac{\partial W_{it}}{\partial BC_{it}} \frac{BC_{it}}{W_{it}}$$

where $c_{it} = \left(\frac{\sum_{\substack{i'=1 \\ i' \neq i}}^{n} W_{i'}}{\sum_{j=1}^{n} W_{jt}}\right) > 0$.

Thus, whether procyclical spending is beneficial only depends on whether $\frac{\partial W_{it}}{\partial BC_{it}} \frac{BC_{it}}{W_{it}}$ is positive.

Recall that $W_{it} = q_{it}^* \times \varepsilon_{q_{it}} \times p_{it}$. Hence we can write $\frac{\partial W_{it}}{\partial BC_{it}} \frac{BC_{it}}{W_{it}}$ as

$$\text{(A.3)} \qquad \frac{\partial W_{it}}{\partial BC_{it}} \frac{BC_{it}}{W_{it}} = \frac{\partial\left(q_{it}^* \times \varepsilon_{q_{it}} \times p_{it}\right)}{\partial BC_{it}} \frac{BC_{it}}{q_{it}^* \times \varepsilon_{q_{it}} \times p_{it}} =$$

$$= \left[\frac{\partial q_{it}^*}{\partial BC_t} \times \varepsilon_{q_{it}} \times p_{it} + \frac{\partial \varepsilon_{q_{it}}}{\partial BC_t} \times q_{it}^* \times p_{it} + \frac{\partial p_{it}}{\partial BC_t} \times q_{it}^* \times \varepsilon_{q_{it}}\right] \left(\frac{BC_{it}}{q_{it}^* \times \varepsilon_{q_{it}} \times p_{it}}\right)$$

$$= \frac{\partial q_{it}^*}{\partial BC_t} \frac{BC_{it}}{q_{it}^*} + \frac{\partial \varepsilon_{q_{it}}}{\partial BC_t} \frac{BC_{it}}{p_{it}} + \frac{\partial p_{it}}{\partial BC_t} \frac{BC_{it}}{p_{it}}$$

$$= \varepsilon_{\text{demand}} + \varepsilon_{\text{mark.eff.}} + \varepsilon_{\text{profit}}.$$

ONLINE APPENDIX B: BAYESIAN ESTIMATION STEPS

This appendix details the estimation steps of the Dynamic Hierarchical Linear Model with Transfer Function [TF-DHLM]. We estimate the model with Bayesian methods (Markov Chain Monte Carlo; MCMC), although we note that frequentist methods could be developed as well. [1]

TF-DHLM

Model. We start with the system of equations for the TF-DHLM (see also equations 17 to 22 in Section 3 of the paper). The dependent variable is indicated by Y_{it}. The model has two dimensions: country *i* (*i* = 1,…,*n*), and period *t* (*t* = 1,…,*T*). All variables with time-varying effects will be named F_{1it} with time-varying parameters θ_{it}. All variables that only have cross-sectional variation in the effects are captured in K_{it}, with parameters γ_i. All variables in the transfer function are called X_{it}, with parameters ψ_i. The system of equations for the TF-DHLM is:

(B.1) $Y_{it} = F_{1it}\theta_{it} + K_{it}\gamma_i + \nu_{1it}$ (Observation Equation)

(B.2) $\theta_{it} = F_{2t}\theta_t + X_{it}\psi_i + \nu_{2it}$ (Structural Equation)

(B.3) $\theta_t = G\theta_{t-1} + \omega_t$ (System Equation)

(B.4) $\gamma_i = \bar{\gamma} + u_{1i}$ (Second-Stage Equation)

(B.5) $\psi_i = \bar{\psi} + u_{2i}$ (Second-Stage Equation)

The mapping function F_{2t} in (B.2) equals 1, mapping the time-varying parameter of each country onto a common time-varying parameter θ_t. In (B.3) the matrix G is set to an identity matrix (West and Harrison 1997), allowing for a random walk in the response parameters, i.e. $\theta_t = \theta_{t-1} + \omega_t$.

The errors of the observation equation and structural equations are heteroskedastic,

(B.6) $\nu_{1it} \sim N\left(0, \sigma^2_{\nu_1,i}\right)$, and $\nu_{2it} \sim N\left(0, \sigma^2_{\nu_2,i}\right)$.

The error variance of each parameter is different. For example, the error for equation (B.4) is different for each cross-sectional variable in the observation equation. We indicate these different parameters by subscripts k_1, k_2, and k_3:[2]

(B.7) $\omega_{t,k_1} \sim N(0, \sigma^2_{\omega,k_1})$, $u_{1i,k_2} \sim N(0, \sigma^2_{u_1,k_2})$, and $u_{2i,k_3} \sim N(0, \sigma^2_{u_2,k_3})$.

---

[1] For the estimation of the DHLM additional details can be found in several articles (Gamerman and Migon 1993; Landim and Gamerman 2000; and Neelamegham and Chintagunta 2004).

[2] In the main text, these different errors per parameter are implied, and for exposition purposes we do not include the subscript itself. In the appendix, we do add this subscript *k*, as it helps in clarifying the sampling scheme.

In the application, the TF-DHLM has two time-varying parameters, the intercept and marketing effectiveness, so k_1 = 1, 2. There are four parameters in γ_i in the observation equation where we shrink the cross-sectional variation, so k_2 = 1,…,4. There are two parameters in ψ_i in the transfer function where we shrink the cross-sectional variation in a second-stage equation: (i) the business-cycle parameter in the transfer function of the marketing sensitivity, and (ii) the economic trend in the growth parameter, so we use k_3 = 1,2.

Now we add equations for the endogenous variable, i.e., the marketing budget for the *n* countries. We estimate the endogenous equations together with the observation equation using the following notation instead of equation (B.1),

(B.8) $$\begin{bmatrix} Y_{it} \\ F_{it}^{Endo} \end{bmatrix} = \begin{bmatrix} F_{it}^{Endo}, F_{it}^{Exo} & 0 \\ 0 & F_{it}^{Exo}, Z_{it}^{IV} \end{bmatrix} \begin{bmatrix} \theta_{it} \\ \gamma_{Endo,1i} \end{bmatrix} + \begin{bmatrix} K_{it}\gamma_i \\ K_{it}\gamma_{Endo,2i} \end{bmatrix} + \begin{bmatrix} \nu_{1it} \\ \nu_{Endo,it} \end{bmatrix},$$

where F_{it}^{Endo} are the endogenous variables. The explanatory variables of the endogenous equations consist of the exogenous control variables of the observation equation (F_{it}^{Exo}), and additional instrumental variables Z_{it}^{IV}. The estimates of the observation equation are corrected for endogeneity by correlating the errors of the observation equation with those of the endogenous equation (Rossi et al., 2005, Section 7.1). We do not adopt a second stage for the parameters in the endogenous equations, and hence they do not have a common hypermean. In the rest of this appendix, we combine the notation of the various γ parameters into $\dot{\gamma}_i = (\gamma_i, \gamma_{Endo,1i}, \gamma_{Endo,2i})$. The hypermean $\bar{\gamma}$ only applies to γ_i, since the parameters of the endogenous equation ($\gamma_{Endo,1i}$, $\gamma_{Endo,2i}$) are not shrunk to a common hypermean.

The error term of (B.8) has a block structure, where the observation equation of country *i* is correlated with the endogenous equation of country *i*, $\begin{bmatrix} \nu_{1it} \\ \nu_{Endo,it} \end{bmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_{\nu,i}^2 = \begin{bmatrix} \sigma_{\nu_1,i}^2 & \sigma_{12} \\ \sigma_{12}' & \sigma_{Endo,i}^2 \end{bmatrix} \right)$.

Estimation. We use MCMC for Bayesian estimation. All steps except the first are iterated for a number of runs. We use 50,000 draws for burn-in and 50,000 draws for inferences. We verified the convergence by plotting and inspecting the posterior parameter draws.

Step 1: Initialization

We begin the iteration scheme with starting values for all parameters, given by: $(\theta_{it}^0, \theta_t^0, \dot{\gamma}_i^0, \bar{\gamma}^0, \psi_i^0, \bar{\psi}^0, (\sigma_{\nu,i}^2)^0, (\sigma_{\nu_2,i}^2)^0, (\sigma_{\omega,k_1}^2)^0, (\sigma_{u_1,k_2}^2)^0, (\sigma_{u_2,k_3}^2)^0)$. As starting values, we use an identity matrix for $(\sigma^2_{v,i})^0$ and $(\sigma^2_{v_2,i})^0$, and a vector of ones for $(\sigma^2_{\omega,k_1})^0, (\sigma^2_{u_1,k_2})^0, (\sigma^2_{u_2,k_3})^0$. For the time-varying parameters we explain the initialization below.

Step 2: Sample Time-Varying Parameters

Step 2a: Forward-filtering backward-sampling algorithm of θ^s_t.

In the *s*-th iteration of the chain, we sample the common underlying time-varying parameter

(B.9) $$\theta^s_t | F_{2t},\ X_{it}, \theta^{s-1}_{it}, \psi^{s-1}_i, (\sigma^2_{v_2,i})^{s-1}, (\sigma^2_{\omega,k_1})^{s-1}.$$

We need to specify a prior for the time-varying parameter at *t*=0. In particular, $\theta^s_0 \sim N(m_0, C_0)$. This prior m_0 should be set in the proximity of the expected value of the time-varying parameter. For the marketing parameter, we use the value .22 reported in the meta-analytic study on the advertising elasticity in the eighties (Assmuss et al. 1984). Given the expected range for the parameter, we set C_0 at 0.1.

Conditional on the right-hand side of (B.9), West and Harrison (1997) derive:

- The *prior* at time *t* is: $\theta^s_t | D_{t-1} \sim N(a_{2t}, R_{2t})$, where $a_{2t} = m_{t-1}$, and $R_{2t} = C_{t-1} + (\sigma^2_{\omega,k_1})^{s-1}$.
- The *one-step-ahead forecast* at time *t* is: $\theta^{s-1}_{it} - X_{it}\psi^{s-1}_i | D_{t-1} \sim N(f_{2t}, Q_{2t})$, where $f_{2t} = F_{2t}a_{2t}$, and $Q_{2t} = F_{2t}R_{2t}F'_{2t} + (\sigma^2_{v_2,i})^{s-1}$. Since the parameters in the transfer function are known in this iteration, the block $X_{it}\psi^{s-1}_i$ is moved to the left-hand side of the equation.
- Then the posterior distribution at time *t* is: $\theta^s_t | D_t \sim N(m_{2t}, C_{2t})$, where $m_{2t} = a_{2t} + R_{2t}F'_{2t}Q^{-1}_{2t}(\theta^{s-1}_{it} - X_{it}\psi^{s-1}_i - f_{2t})$, and $C_{2t} = R_{2t} - R_{2t}F'_{2t}Q^{-1}_{2t}F_{2t}R'_{2t}$.
- To sample from this conditional distribution we adopt the forward-filtering backward-sampling algorithm (Carter and Kohn 1994; Frühwirth-Schnatter 1994).
    - For $t = 1,\ldots,T$ we apply forward filtering to obtain the moments, i.e., m_{2t} and C_{2t}. Note that for $t = 0$ we use the prior information, m_0 and C_0.
    - At $t = T$ we sample a vector of system parameters from distribution $N(m_{2t}, C_{2t})$.
    - Then we sequence backward for $t = T\text{-}1, T\text{-}2, \ldots, 1$ sampling from: $p(\theta^s_t | \theta^s_{t+1}) \sim N(q^*_{2t}, Q^*_{2t})$, where $q^*_{2t} = m_{2t} + B_t(\theta^s_{t+1} - a_{2t+1})$, $Q^*_{2t} = C_{2t} - B_tR_{2t+1}B'_t$, and $B_t = C_{2t}R^{-1}_{2t+1}$.

Step 2b: Mapping

To sample the *s*-th iteration of θ_{it}^{s}, we adopt the sequential inference step from Neelemaghan and Chintagunta (2004). The prior distribution is:

(B.10) $\theta_{it}^{s} | \tilde{Y}_{it}, F_{1it}, F_{2t}, X_{it}, \theta_{t}^{s}, K_{it}, \dot{\gamma}_{i}^{s-1}, \psi_{i}^{s-1}, \left(\widetilde{\sigma_{\nu,\iota}^{2}}\right)^{s-1}, \left(\sigma_{\nu_2,i}^{2}\right)^{s-1} \sim N(a_{1it}, R_{1it})$

The superscript of θ_{t}^{s} is now "*s*", since this has been sampled in iteration step 2a. The tilde"~" on Y and $\sigma_{\nu,i}^{2}$ indicates the mean and variance of the observation equation, conditional on the endogenous equation. The conditional mean is $\tilde{Y}_{it} = \ (Y_{it} - K_{it}\gamma_{i}^{s-1}) - \sigma_{12}\left(\sigma_{Endo,i}^{2}\right)^{-1}(F_{it}^{endo} - K_{it}\gamma_{Endo,2i}^{s-1} - Z_{it}\gamma_{Endo,1i}^{s-1})$, and the conditional variance is $\widetilde{\sigma_{\nu,\iota}^{2}} = \sigma_{\nu_1,i}^{2} - \sigma_{12}\left(\sigma_{Endo,i}^{2}\right)^{-1}\sigma_{12}'$.

Now we can specify the *prior* distribution, with mean $a_{1it} = F_{2t}\theta_{t}^{s} + X_{it}\psi_{i}^{s-1}$ and variance $R_{1it} = \left(\sigma_{\nu_2,i}^{2}\right)^{s-1}$. These are used for the one-step-ahead forecast, $\tilde{Y}_{it}|D_{t-1} \sim N(f_{1it}, Q_{1it})$, where $f_{1it} = F_{1it}a_{1it}$, and $Q_{1it} = F_{1it}R_{1it}F_{1it}' + \left(\widetilde{\sigma_{\nu,\iota}^{2}}\right)^{s-1}$. The updated posterior is $\theta_{it}^{s}|D_{t} \sim N(m_{1it}, C_{1it})$, with $m_{1it} = a_{1it} + R_{1it}F_{1it}'Q_{1it}^{-1}(\tilde{Y}_{it} - f_{1it})$, and $C_{1it} = R_{1it} - R_{1it}F_{1it}'Q_{1it}^{-1}F_{1it}R_{1it}'$.

Step 3: Sample $\boldsymbol{\sigma_{\nu,i}^{2}}$

The *s*-th iteration of the covariance matrix of the observation equation is sampled by

(B.11) $\left(\sigma_{\nu,i}^{2}\right)^{s} | Y_{it}, F_{1it}, K_{it}, Z_{it}^{IV}, \theta_{it}^{s}, \dot{\gamma}_{i}^{s-1}$

We sample the 2 by 2 matrices $\left(\sigma_{\nu,i}^{2}\right)^{s}$ from an Inverse Wishart distribution, sampling the error of the observation equation and the endogenous equation for each country *i*. We specify a diffuse prior $IW\left(v_{\nu}^{prior}, S_{\nu}^{prior}\right)$, where $v_{\nu}^{prior} = 2$, and $S_{\nu}^{prior} = 0.01 * I_2$. The conditional posterior distribution then becomes

$$IW\left(v_{\nu}^{prior} + T, S_{\nu}^{prior} + \sum_t \begin{bmatrix} Y_{it} - (F_{1it}\theta_{it}^{s} + K_{it}\gamma_{i}^{s-1}) \\ F_{it}^{Endo} - (Z_{it}^{IV}\gamma_{Endo,1i}^{s-1} + K_{it}\gamma_{Endo,2i}^{s-1}) \end{bmatrix} * \begin{bmatrix} Y_{it} - \left(F_{1it}\theta_{it}^{s} + K_{it}\gamma_{i}^{s-1}\right) \\ F_{it}^{Endo} - (Z_{it}^{IV}\gamma_{Endo,1i}^{s-1} + K_{it}\gamma_{Endo,2i}^{s-1}) \end{bmatrix}'\right).$$

Step 4: Sample $\boldsymbol{\sigma^2_{v_2,i}}$

$\sigma^2_{v_2,i}$ is a diagonal matrix of heteroskedastic variances for the structural equations. In the application, we have 36 structural equations: 18 countries with a time-varying growth parameter and a time-varying marketing effect. The *s*-th iteration is:

(B.12) $\left(\sigma^2_{v_2,i}\right)^s | \, F_{2t}, X_{it}, \; \theta^s_{it}, \theta^s_t, \psi^{s-1}_i.$

We sample the variance from an Inverse Gamma distribution. We use a diffuse prior, $IG\left(\frac{v^{prior}_{v_2}}{2}, \frac{S^{prior^{-1}}_{v_2}}{2}\right)$, where $v^{prior}_{v_2} = 1$, and $S^{prior}_{v_2} = 0.001$. The posterior is: $IG\left(\frac{v^{prior}_{v_2}+T}{2}, \frac{(S^{prior}_{v_2}+\sum_t\left(\theta^s_{it}-F_{2t}\theta^s_t-X_{it}\psi^{s-1}_i\right)^2)^{-1}}{2}\right).$

Step 5: Sample $\boldsymbol{\sigma^2_\omega}$

The sampling of the elements of σ^2_ω is separate for each time-varying parameter in the observation equation, indicated by k_1:

(B.13) $\left(\sigma^2_{\omega,k_1}\right)^s | \, \theta^s_t$

We sample this from an Inverse Gamma distribution with prior, $IG\left(\frac{v^{prior}_{\omega}}{2}, \frac{S^{prior^{-1}}_{\omega}}{2}\right)$, where $v^{prior}_\omega = 1$, and $S^{prior}_\omega = 0.001$. The conditional posterior distribution is $IG\left(\frac{v^{prior}_{\omega}+T}{2}, \frac{(S^{prior}_{\omega}+\sum_t(\theta^s_t-\theta^s_{t-1})^2)^{-1}}{2}\right).$

Step 6: Cross-Sectional Parameters of the Observation Equation

The cross-sectional parameters of the observation equation are allowed to differ over countries as per equation (B.4).

<u>*Step 6a*</u>: Sample $\dot{\gamma}$ (including endogeneity parameters)

The parameters in $\dot{\gamma}$ consist of three parts (see below equation B.8), that can be sampled together: (i) the control variables from the observation equation (γ_i), (ii) the control variables in the endogenous equation ($\gamma_{Endo,2i}$), and (iii) the parameters of the instrumental variables in the endogenous equation ($\gamma_{Endo,1i}$). These three parts can be sampled together, using the block structure of a Seemingly Unrelated Regression (SUR). The *s*-th iteration is sampled by

(B.14) $\dot{\gamma}^s_i | Y_{it}, \; F_{1it}, K_{it}, Z^{IV}_{it}, \theta^s_{it}, \bar{\gamma}^{s-1}, \left(\sigma^2_{v,i}\right)^s, \left(\sigma^2_{u_1,k_2}\right)^{s-1}.$

The difference between the three parts of $\dot{\gamma}$ is that the two parts from the endogenous equation are not shrunk to a hypermean, so we use a diffuse normal prior for these parameters, $N\left(\mu_{\gamma^{\text{endo}}}^{prior}, \sigma_{\gamma^{\text{endo}}}^{2,prior}\right)$, where $\mu_{\gamma^{\text{endo}}}^{prior} = 0$, and $\sigma_{\gamma^{\text{endo}}}^{2,prior} = 1000$. The prior of the parameters in the observation equation come from a common distribution, $N(\bar{\gamma}_{k_2}^{s-1}, \left(\sigma_{u_1,k_2}^2\right)^{s-1})$.

The dependent variables of all equations are stacked in $\boldsymbol{Y}$, which is a 2*T*n-by-1 vector, $\boldsymbol{Y} = \begin{pmatrix} Y_{it} - (F_{1it}\theta_{it}^s) \\ F_{it}^{Endo} \end{pmatrix}$. F_{it}^{Endo} is the endogenous variable. Since in this iteration time-varying parameters are known, they are part of the left-hand side. The independent variables are combined in a block structure $\boldsymbol{K} = \begin{bmatrix} K_{it} \otimes I_n & \boldsymbol{0} \\ \boldsymbol{0} & (Z_{it}^{IV}, K_{it}) \otimes I_n \end{bmatrix}$. Since some of the equations that form vector $\boldsymbol{Y}$ and matrix $\boldsymbol{K}$ are correlated, we have to transform them by using the root of the cross equation covariance matrix, which is matrix U_1 in $(\sigma_v^2)^s = Cov(v) = U_1' U_1$. We now create $\widetilde{\boldsymbol{Y}}$ and $\widetilde{\boldsymbol{K}}$ of an uncorrelated system of equations, with $\widetilde{\boldsymbol{Y}} = (U_1^{-1} \otimes I_T)\boldsymbol{Y}$ and $\widetilde{\boldsymbol{K}} = (U_1^{-1} \otimes I_T)\boldsymbol{K}$.

We now sample $\dot{\gamma}$ from the following multivariate normal distribution $N\left(\left(\widetilde{\boldsymbol{K}}'\widetilde{\boldsymbol{K}} + A^{-1}\right)^{-1}\left(\widetilde{\boldsymbol{K}}'\widetilde{\boldsymbol{Y}} + A^{-1} * m\right), \left(\widetilde{\boldsymbol{K}}'\widetilde{\boldsymbol{K}} + A^{-1}\right)^{-1}\right)$, where A is a diagonal matrix with the variance of the prior distribution (i.e., σ_{u_1,k_2}^2 for the observation equations, and the diffuse prior for the endogenous equation), similarly m is a vector of the mean of the prior distributions.

<u>*Step 6b*</u>: Sample $\bar{\gamma}$

For the s-th iteration of the hypermean of the time-invariant parameters we use:

(B.15) $\qquad \bar{\gamma}^s \,|\, \gamma_i^s, \left(\sigma_{u_1,k_2}^2\right)^{s-1}$.

Note that on the right-hand side we use γ_i^s, instead of $\dot{\gamma}_i^s$, since we only use shrinkage on the cross-sectional parameters of the observation equation and not of the endogenous equation. This is a Gibbs sampling step from a normal distribution, with a diffuse normal prior where $\mu_{\bar{\gamma}}^{prior} = 0$, and $\sigma_{\bar{\gamma}}^{2,prior} = 1000$. We sample the hypermean for each variable k_2 from the posterior distribution,

$N\left(\left(\iota_n'\iota_n + \left(\sigma_{\bar{\gamma}}^{2,prior}\right)^{-1}\right)^{-1}\left(\iota_n' vec(\gamma_{i,k_2}^s) + \left(\sigma_{\bar{\gamma}}^{2,prior}\right)^{-1}\mu_{\bar{\gamma}}^{prior}\right), \left(\sigma_{u_1,k_2}^2\right)^{s-1} * \left(\iota_n'\iota_n + \left(\sigma_{\bar{\gamma}}^{2,prior}\right)^{-1}\right)^{-1}\right)$,

where ι_n is a unit vector of size n.

<u>*Step 6c*</u>: Sample σ_{u_1,k_2}^2

The final part of the cross-sectional variables in the observation equation is the error variance of the shrinkage equations:

(B.16) $\left(\sigma^2_{u_1,k_2}\right)^s |\, \gamma_i^s, \bar{\gamma}^s$

We sample this parameter for each variable separately from an Inverse Gamma distribution. We use a diffuse Inverse Gamma prior, where $v_{u_1}^{prior} = 1$, and $S_{u_1}^{prior} = 0.01$. The conditional posterior distribution is $IG\left(\frac{v_{u_1}^{prior}+n}{2}, \frac{(S_{u_1}^{prior}+\sum_i\left(\gamma_i^{s,k_2}-\bar{\gamma}^{s,k_2}\right)^2)^{-1}}{2}\right)$.

Step 7: Cross-sectional Parameters of Structural Equation

For the parameters in ψ_i we distinguish between the parameters in the structural equations that are shrunk in a second stage and those that are not. The fixed effects ($\psi_{0\mathrm{i}}$) in the equations for the growth parameter are estimated separately, whereas the parameter of the economic trend ($\psi_{1\mathrm{i}}$) in the equations of the growth parameter, and the effects of the business cycle ($\psi_{2\mathrm{i}}$) in the equations of the marketing sensitivity have a common hypermean. For steps 7a and 7b we use a block structure as in step 6a, even though this is not directly necessary in our case since we have a diagonal covariance matrix, and hence the dependent variables are not correlated. The reason to write it as a SUR model specification is that it makes this sampling scheme easy to extend to a model with full covariance matrix if required.

Step 7a: Sample ψ_i

The s-th iteration of ψ_i is sampled by:

(B.17) $\psi_i^s |\, F_{2t}, \theta_{it}^s, \theta_t^s, \bar{\psi}^{s-1}, \left(\sigma^2_{\nu_2,i}\right)^s, \left(\sigma^2_{u_2,k_3}\right)^{s-1}$

As said, ψ_i consists of three parts where the part of the fixed effects ψ_{0i} is not pooled; for this part we use a diffuse normal prior, where $\mu_{\psi_0}^{prior} = 0$, and $\sigma_{\psi_0}^{2,prior} = 1000$. The prior of the other two parts comes from the common distribution $N(\bar{\psi}_{k_3}^{s-1}, \left(\sigma^2_{u_2,k_3}\right)^{s-1})$.

The dependent variable is a stacked vector of all the time-varying parameters of the observation equation, $\mathrm{vec}(\theta_{it}^s - F_{2t}\theta_t^s)$, which is a T*n-by-1 vector. Note that we moved the part of the right-hand side of the structural equation that is known in this iteration (i.e., $F_{2t}\theta_t^s$) to the left-hand side. The block structure of the independent variables is created by the Kronecker product of the variables of the structural equation and the identity matrix of size n, $\boldsymbol{X} = X_{it} \otimes I_n$. U_2 from $\left(\sigma^2_{\nu_2}\right)^s = Cov(\nu_2) = U_2' U_2$ creates an uncorrelated system of equations:

$\widetilde{\boldsymbol{\theta}} = (U_2^{-1} \otimes I_T)\text{vec}(\theta_{it}^s - F_{2t}\theta_t^s)$, and $\widetilde{\boldsymbol{X}} = (U_2^{-1} \otimes I_T)\boldsymbol{X}$. Now ψ_i is sampled from its multivariate posterior distribution, $N\left(\left(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}} + A\right)^{-1}\left(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{\theta}} + A * m\right), \left(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}} + A\right)^{-1}\right)$, where *A* is a diagonal matrix with the variance of the prior distribution (i.e., σ_{u_2,k_3}^2), and *m* is a vector with the means of the prior distributions.

<u>*Step 7b*</u>: Sample $\bar{\psi}$

This step only applies to $\psi_{1\text{i}}$ and $\psi_{2\text{i}}$, since $\psi_{0\text{i}}$ is not pooled. The *s*-th iteration of the hypermean of the time-invariant parameters is given by:

(B.18) $\bar{\psi}^s |\, \psi_i^s, \left(\sigma_{u_2,k_3}^2\right)^{s-1}$.

This is a Gibbs sampling step, with a diffuse normal prior with $\mu_{\bar{\psi}}^{prior} = 0$, and $\sigma_{\bar{\psi}}^{2,prior}$= 1000. We sample the hypermean for each variable k_3 from its posterior,

$$N\left(\left(\iota_n'\iota_n + \left(\sigma_{\bar{\psi}}^{2,prior}\right)^{-1}\right)^{-1}\left(\iota_n' vec(\psi_{i,k_3}^s) + \left(\sigma_{\bar{\psi}}^{2,prior}\right)^{-1}\mu_{\bar{\psi}}^{prior}\right), \left(\sigma_{u_2,k_3}^2\right)^{s-1} * \left(\iota_n'\iota_n + \left(\sigma_{\bar{\psi}}^{2,prior}\right)^{-1}\right)^{-1}\right),$$

where ι_n is a unit vector of size *n*.

<u>*Step 7c*</u>: Sample $\sigma_{u_2,3}^2$

The final part of the variables in the structural equation is the error of the shrinkage equations, which again only applies to the parameters with a second-stage equation;

(B.19) $\left(\sigma_{u_2,k_3}^2\right)^s |\, \psi_i^s, \bar{\psi}^s$

We sample this parameter for each variable separately from an Inverse Gamma distribution. We use a diffuse prior with $v_{u_2}^{prior} = 1$, and $S_{u_2}^{prior} = 0.01$. The conditional posterior distribution is

$$IG\left(\frac{v_{u_2}^{prior} + n}{2}, \frac{\left(S_{u_2}^{prior} + \Sigma_i\left(\psi_i^{s,k_3} - \bar{\psi}^{s,k_3}\right)^2\right)^{-1}}{2}\right).$$

Iterating By iterating through steps 2 to 7 we sample the parameters and other model statistics of the TF-DHLM. We verified the correctness of the estimation code through extensive simulations. The Matlab code is available on request from the first author.

ONLINE APPENDIX C:

Table C1: Annual Absolute Number of Holiday Visitors to New Zealand (1981-2011)

	Mean (s.d.)	Minimum	Maximum
Australia	207,334 (113,977)	87,801	475,715
United States	105,590 (25,084)	52,794	136,378
Japan	92,082 (40,455)	17,608	141,815
United Kingdom	72,125 (52,255)	9,628	167,286
South Korea	36,376 (35,116)	72	111,547
Germany	30,811 (15,467)	5,908	51,019
Canada	20,024 (5,332)	11,500	30,455
China	19,793 (28,229)	176	88,765
Taiwan	17,710 (14,262)	924	50,721
Singapore	14,456 (6,255)	2,324	25,097
Hong Kong	11,372 (6,275)	1,509	21,270
Thailand	9,140 (7,130)	340	25,308
The Netherlands	9,042 (6,673)	1,432	18,937
Malaysia	8,284 (4,739)	1,204	16,513
Switzerland	7,898 (2,864)	1,756	11,557
Sweden	5,784 (2,634)	956	9,529
France	5,086 (4,535)	732	15,818
Indonesia	4,277 (2,511)	996	9,835

Source: http://www.stats.govt.nz/.

Table C2: Correlations between Model Variables of the Demand Model

	$\Delta \ln q_{it}$	$\Delta \ln M_{it}$	$\Delta \ln M_{it} * \ln BC_{it}$	$\Delta \ln BC_{it}$	$\Delta \ln Price_{it}$	$\Delta \ln OilPr_t$	$\Delta \ln Flights_{it}$	$\Delta \ln EconTrend_{it}$
$\Delta \ln q_{it}$	1	0.14***	-0.18***	0.10**	-0.09**	-0.09**	0.17***	0.23***
$\Delta \ln M_{it}$		1	0.28***	0.00	0.08*	-0.02	-0.07	0.13***
$\Delta \ln M_{it} * \ln BC_{it}$			1	-0.15***	-0.01	0.00	0.03	-0.04
$\Delta \ln BC_{it}$				1	0.02	0.35***	0.12***	-0.00
$\Delta \ln Price_{it}$					1	0.11***	0.01	0.04
$\Delta \ln OilPr_t$						1	0.05	-0.07*
$\Delta \ln Flights_{it}$							1	0.10
$\Delta \ln EconTrend_{it}$								1

*, **, *** indicate significance at the 10%, 5%, and 1% significance level

Table C3: Correlations between Model Variables of the Profit Contribution Model

	$\ln p_{it}$	$\Delta \ln BC_{it}$	$\ln EconTrend_{it}$	$\ln Price_{it}$	$\ln OilPr_t$
$\ln p_{it}$	1	-0.02	0.22**	-0.30***	-0.01
$\Delta \ln BC_{it}$		1	-0.02	0.01	0.27***
$\ln EconTrend_{it}$			1	-0.13	0.14
$\ln Price_{it}$				1	0.03
$\ln OilPr_t$					1

*, **, *** indicate significance at the 10%, 5%, and 1% significance level

ONLINE APPENDIX D: ADDITIONAL ROBUSTNESS CHECKS

Table D1 shows a model (BM6) where we refrain from using extrapolated data for regional marketing spend (see explanation footnote b in Table 2), effectively omitting the data from the 1981-1987 era. The sign and significance of the effects of the independent variables remain the same as for the focal model.

Table D1: Estimation Results Demand Model and Robustness Checks

	Expectation	TF-DHLM	BM6 No extrapolated regional marketing spend
Key Response Parameters			
Tourism Marketing Elasticity [a]	+	0.14***	0.08***
Business Cycle Elasticity	+	0.54**	0.43**
Moderating Effect of Business Cycle on Tourism Marketing Elasticity	+/-	-2.07***	-2.00**
Response Parameters in Growth Function			
Elasticity to Economic Growth	+	2.94***	3.09***
Response Parameters for Control Variables			
Relative Price Elasticity	-	-0.16**	-0.16**
Oil Price Elasticity	-	-0.02	0.04
Number of Flights Offered	+	0.14***	0.23***

*, **, *** indicate significance at the 10%, 5%, and 1% significance level based on the Highest Posterior Density interval. This interval is tested one-sided when we expect a specific direction of the effect, and two-sided otherwise (i.e., for the interaction variable).

a. Mean across time

We have also expanded the focal model with the interaction between oil price and the business cycle, as well as the interactions of the business cycle with the other control variables. The results are given below. These results show that none of the interactions are significant. To save space, we do not include these extensions in the paper.

Table D2: Estimation Results Demand Model Compared to Robustness Check with Additional Interactions

	Expectation	TF-DHLM	Interactions of Control Variables With Business Cycle
Key Response Parameters			
Tourism Marketing Elasticity [a]	+	0.14***	0.04
Business Cycle Elasticity	+	0.54**	0.53**
Moderating Effect of Business Cycle on Tourism Marketing Elasticity	+/-	-2.07***	-2.39***
Response parameters in Growth Function			
Elasticity to Economic Growth	+	2.94***	3.26***
Response Parameters for Control Variables			
Relative Price Elasticity	-	-0.16**	-0.13**
Oil Price Elasticity	-	-0.02	-0.01
Number of Flights Offered	+	0.14***	0.13***
Interaction of the Business Cycle and the Relative Price Elasticity	+/-		0.94
Interaction of the Business Cycle and the Oil Price Elasticity	+/-		0.06
Interaction of the Business Cycle and the Number of Flights Offered	+/-		0.22

*, **, *** indicate significance at the 10%, 5%, and 1% significance level based on the Highest Posterior Density interval. This interval is tested one-sided when we expect a specific direction of the effect, and two-sided otherwise (i.e., for the interaction variables).

a. Mean across time

ONLINE APPENDIX E: OPTIMAL BUDGET ALLOCATION

Allocation Rules

In the application, demand is measured by the number of tourists from country *i* in period *t*: q_{it}, whereas the profit contribution is p_{it}. The marketing budget is the marketing expenditures in period *t* (in NZ dollars) for region *r*: M_{rt}. One difference with the set-up of Fisher et al. (2011) is that we study the allocation of the budget to region *r* rather than country *i*. This is in line with the institutional settings observed in the data, where marketing budgets are determined per region and not per country. M_t is the total budget for marketing available in period *t*, as set exogenously by the government.

As such, we optimize the profit Π_t across the 18 countries by setting the optimal budget in the five regions:

(D.1) $$\max_{\mathrm{M_{rt}}} \Pi_{\mathrm{t}} = \sum_{i=1}^{18} p_{it} \times q_{it} - M_t,$$

subject to

(D.2) $$q_{it} = g_i(t) \times f_i(BC_{it}, M_{it}, H_{it}).$$

(D.3) $$\sum_{r=1}^{5} M_{rt} \leq M_t.$$

Fischer et al. (2011) derive both the theoretically optimal allocation (their equations 5 and 6) and a heuristic that is near optimal (their equations 9 and 10). An issue with the theoretically optimal allocation is that it is an internal-point solution, requiring numerical optimization (Fischer et al. 2011). The heuristic overcomes this issue, and gives a closed-form solution, which enhances its usability (Albers 2012). Therefore, we use the heuristic,[3] adapted to our setting where the budget is set per region. The optimal budget to allocate to region *r* is based on the following equation:

(D.4) $$M_{rt}^{opt} = s_{rt} * M_t = \frac{\sum_{i \in C_r} W_{it}}{\sum_{j=1}^{18} W_{jt}} * M_t,$$

where C_r is the subset of countries in region *r* (e.g., Canada and the U.S. in North America). W_{it} is obtained by:

(D.5) $$W_{it} = \hat{q}_{it} \times \hat{\theta}_{1it} \times \hat{p}_{it},$$

where $\hat{\theta}_{1it}$ is the estimated marketing elasticity from Eq. 12 and $\hat{q}_{it}$ is the expected number of visitors, which we calculate by multiplying the observed number of visitors in the period before, q_{it-1}, by the estimated growth multiplier, which we define as $\hat{\rho}_{it}$. For the growth multiplier, we

[3] We also verified the allocation prescribed when using exact numerical optimization, and obtained very similar outcomes (available on request from the first author).

capitalize on the fact that we have a model in first differences, where the intercept estimate ($\hat{\theta}_{0it}$) from Eq. 12 is the (ceteris paribus) growth rate in the log number of visitors (see also Lamey et al. 2007). Hence, the growth multiplier for the number of visitors is given by $\hat{\rho}_{it} = \exp(\hat{\theta}_{0it})$. For both estimates, we first take the median across draws. In case *t* refers to multiple years, we first calculate the weight per year and then take the average.[4]

For $\hat{p}_{it}$, we use the expectation based on model (23) for the eight countries for which we have data on average expenditures per visitor from the New Zealand Ministry of Innovation, Business and Employment in their International Visitors Survey (http://www.med.govt.nz/sectors-industries/tourism/tourism-research-data/international-visitor-survey/online-database). These data are not available for all countries, and hence missing average expenditures for these countries (e.g., the Netherlands) are approximated by the $\hat{p}_{it}$ from similar countries in the region (e.g., the average across Germany and the U.K.).

For each allocation, we calculate expected demand. To ascertain the stability of the results, we use 5,000 posterior parameter draws. That is, for each allocation we obtain 5,000 estimates for expected demand. This allows us to compute the likelihood that one type of allocation (A) is better than another (B) by calculating for which fraction of these 5,000 draws allocation A is superior to allocation B. These likelihoods are reported in the last column of Table 7.

References Used in the Appendices

Albers S (2012) Optimizable and implementable aggregate response modeling for marketing decision support. *Int. J. Res. Marketing* 29(2):111–122.

Assmus G, Farley JU, Lehmann DR (1984) How advertising affects sales: Meta-analysis of econometric results. *J. Marketing Res.* 21(1):65–74.

Carter C, Kohn R (1994) On Gibbs sampling for state space models. *Biometrika* 81(3):541–553.

Frühwirth-Schnatter S (1994) Data augmentation and dynamic linear models. *J. Time Series Anal.* 15(2):183–202.

Fischer M, Albers S, Wagner N, Frie M (2011) Dynamic marketing budget allocation across countries, products, and marketing activities. *Marketing Sci.* 30(4):568–585.

Gamerman D, Migon HS (1993) Dynamic hierarchical models. *J. Royal Stat. Soc. Series B (Methodological)* 55(3):629–642.

Ingene CA, Parry ME (1995) A note on multi-regional marketing. *Management Sci.* 41(7):1194-1201.

[4] In some extreme cases, there are some marketing elasticities below zero. If we were to optimize per country, it is optimal to allocate no budget to the countries with negative elasticities. However, since we maximize the budget per region and not per country, Ingene and Parry (1995) conclude that the elasticity may be negative in some individual countries.

Landim F, Gamerman D (2000) Dynamic hierarchical models: An extension to matrix-variate observations. *Comput. Statist. Data Anal*. 35(1):11–42.
Neelamegham R, Chintagunta PK (2004) Modeling and forecasting the sales of technology products. *Quant. Marketing Econom.* 2(3):195–232.
Rossi PE, Allenby GM, McCulloch R (2005) *Bayesian Statistics and Marketing.* Wiley and Sons, New York.
West M, Harrison J (1997) *Bayesian Forecasting and Dynamic Models,* 2nd ed. Springer-Verlag, New York.