

Bayesian Dynamic Panel Models for Tourism Research

Abstract

This paper introduces several innovative dynamic panel data models that allow variations in slope coefficients both across time and cross-sectional units. We replace time variation with a dynamic (autoregressive) component, and introduce several variations of the so-called Mundlak device in which random intercepts are linear function of the average values of the regressors. We develop all our models in a Bayesian framework, and test their performance using an interesting application on the impact of advertising on firm sales. We provide technical details of all these models and present tools to compare their performance in a Bayesian framework. Moreover, model averaging and posterior model pools are presented to gain more insight into the relationship between advertising and sales.

1. Introduction

The use of panel data in tourism research is quite prevalent as panel data allow more precise estimation of regression parameters (Garín-Mun, 2006; Sequeira and Maças Nunes, 2008; Falk, 2010; Rey et al. 2011). However, most applications, in general, have been static in that they do not allow the dependent variable to depend on its past realization. Dynamic panel models (DPM), on the other hand, despite their popularity in other related fields have not been used much in tourism research. They can bring many advantages to several tourism contexts. For example: demand modeling; where “if the impact of past tourism is neglected, the effect of the relevant variable considered will tend to be overestimated (as the estimated coefficients will involve direct and indirect effects)” (Garín-Mun, 2006, p.286). Even in contexts where the lagged dependent variable is not of high interest, the introduction of these lags is essential to control for the dynamic nature of the industry. Adopting the appropriate behavioral specification allow us to control for new or different paths between the dependent and independent variables (Bond, 2002). In addition, dynamic models are highly effective to deal with endogeneity-related issues (i.e. simultaneity, omitted variable bias, country-specific effects and measurement error).

Our aim in this paper is to take the estimation of DPM to a different level, introducing seven different advanced formulations that can be used in future tourism applications on panel data. We focus specifically on the issue of heterogeneity, where most existing panel data models (including DPM) assume that slope coefficients are common across units and heterogeneity is modeled using fixed or random effects for the intercept. However, heterogeneity is not exhausted by modeling the intercepts as unit-specific and using fixed or random effect estimators. Heterogeneity may also be reflected in differences in the slope coefficients. We also argue that heterogeneity in slope coefficients is not necessarily exhausted by assuming that they are random and follow, for example, a multivariate normal distribution. Instead the

slope coefficients may also vary over time, in which case we need models that allow for variation over both time and cross-sectional units. In this respect, dynamic slope coefficient models may be necessary. This is particularly true when the objective is to use panel data models for forecasting purposes, as is common in applied tourism research. For instance, when one is using dynamic panel data models for forecasting, adjustment costs are thought to be important. Such models can be improved, in terms of forecasting, when coefficients are dynamically varying and/or when they are different for each cross-section (using a random-coefficients framework). However, we are not aware of research that accounts for these concerns.

In addition, different specifications for the variation of coefficients across units or over time are rarely tested against alternatives. For example, coefficients can include two components - one that varies over time and one that varies across individuals. An alternative specification is to replace time variation with a dynamic (autoregressive) component. We also discuss the issue of independence between the random intercepts and the regressors, where in simple panel data models, this is the typical assumption. This assumption cannot however always be correct and the fixed effects specification is considered better in this regard. An intermediate assumption, which we present, is to use the so-called Mundlak device in which random intercepts are a linear function of the average values of the regressors. Clearly, this can be extended to the slope coefficients, but this line of research has not yet been pursued in theoretical or applied research.

Finally, we develop all of our models in a Bayesian framework as the traditional sampling – theory estimators are more problematic to compute and they may behave erratically in finite samples. With dynamic panel data models with random coefficients, in particular, the use of sampling-theory estimators is prohibited as the widely used Arellano-Bond Generalized Method of Moments (GMM) estimator assumes fixed slope coefficients. As stated by Assaf and Tsionas (2019, p. 273), “this is an important handicap which limits the scope of sampling-theory estimators in dynamic panel data models. Bayesian procedures are more straightforward to apply in dynamic models as lagged dependent variables do not create new problems in terms of estimation for the Bayesian approach”. Besides, use of the Arellano-Bond GMM estimator is problematic when the Arellano-Bond instruments are weak –in which case the sampling behavior of GMM can be erratic and unreliable. The standard criticism against the Bayesian approach in this context (namely that random parameters are orthogonal to the regressors) can be alleviated through the use of the Mundlak device. Use of this device in the context of dynamic panel data is, however, novel and constitutes a major contribution of the present paper.

2. Dynamic Panel Models in Tourism

Seetaram and Petit (2012) provided an interesting and comprehensive review on the application of panel data analysis in tourism research. Generally, and in line with their findings, most panel data applications in tourism continue to focus on either the determinants of tourism demand or the relationship between tourism and economic growth (Yang, 2012; Zhang and Gao, 2016; Du et al. 2016; Bilen et al. 2017; Paramati et al. 2017; Saha et al. 2017; Wu and Wu, 2019). In addition, over the last decade, the tourism literature (and similarly the hotel literature) has experienced an increased number of applications of dynamic panel models (Garín-Mun, 2006; Sequeira and Maças Nunes, 2008; Rey et al., 2011; Seetanah, 2011; Yang, 2012; Assaf et al., 2015; Kim et al., 2019; Woo et al., 2019). Apart from the many advantages in estimation that dynamic models introduce over static models (e.g. fixed effect, random effect, Ordinary Least Squares (OLS)), they also provide a more realistic representation of most modelling contexts in tourism and hospitality; such as the context of demand modeling, where it is more logical to assume that current visitation depends on past visitation (Seetaram and Petit, 2012). Ignoring such dynamic effects may result in an overestimation of the model parameters (Garín-Mun, 2006, p.288).

Three trends can be observed from the existing dynamic panel studies in tourism and hospitality research:

1. First, the use of the Arellano-Bond (AB) technique seems to be the most common approach for estimating dynamic panel models and for deriving both short-run and long-run relationships (Maloney et al. 2005; Naudé and Saayman, 2005; Fayissa et al. 2008; Brida and Risso; Seetanah, 2011; Santana-Gallego et al. 2011; Massidda and Etzo, 2012; Li et al. 2015; Li et al. 2017). The same conclusion was also reported by Seetaram and Petit (2012).
2. Second, given the heavy use of the (AB) technique, most estimations are based on the GMM procedure. Our review of the literature clearly indicates that the Bayesian approach has been absent from most existing studies in tourism and hospitality. As mentioned above and later in the paper, the Bayesian approach introduces clear advantages in the estimation of panel models such as the ability to perform better when T is small and the higher flexibility it offers for complex model specifications such as the ones we are proposing in this study. The Bayesian approach also does not require instrumental variables, as is the case with GMM. When the Arellano-Bond instruments are weak, the sampling behavior of GMM can result in inconsistent estimates.
3. Third, none of the existing studies has fully explored heterogeneity. While accounting for dynamic effects is a critical issue, we argue that there is a far more important issue, namely whether one can reasonably assume that only intercepts are unit-specific, but otherwise the slope coefficients are the same. For many reasons, the latter is,

apparently, true as not all units, at a given point in time, have exactly the same “technology” of transforming explanatory variables to dependent variables (whatever specific meaning this acquires in any given economic narrative). Panel data specifications with slope coefficients that are allowed to differ across cross-sectional units and/or over time have received practically no attention in the literature. The increased use of Arellano and Bover and Arellano and Bond specifications means that researchers become increasingly aware that processes in tourism economics are inherently dynamic (due to, for example, adjustment costs or other inertia). This opens up novel questions. First, are the dynamic responses the same across cross-sectional units? Second, are responses of the dependent variable to explanatory variables time-invariant and/or homogenous across cross-sectional units?

3. Why Bayesian?

In addition to the many advantages of the Bayesian approach that we discussed above, here we provide additional background as to why the Bayesian estimator is highly effective in the context of panel data. Suppose we have a standard panel data autoregressive model of the form:

$$y_{it} = \alpha_i + \rho y_{i,t-1} + u_{it}, i = 1, \dots, n, t = 1, \dots, T, \quad (1)$$

where α_i s are fixed effects. If T tends to infinity, then the least-squares-dummy-variables (LSDV) estimator of α_i and γ are consistent. However, for finite T , and $|\gamma| < 1$, as N tends to infinity it is known that the LSDV estimator of γ is inconsistent due to the incidental parameters problems (Neyman and Scott, 1948). While the maximum likelihood estimators are inconsistent, there have been proposed consistent instrumental variables/methods of moments estimators (e.g. Anderson and Hsiao, 1981, 1982; Arellano and Bond, 1991; Arellano and Bover, 1995; Ahn and Schmidt, 1995; Blundell and Bond, 1998).

If we take first differences, we can eliminate the fixed effects, since:

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta u_{it}, i = 1, \dots, n, t = 1, \dots, T. \quad (2)$$

If $|\gamma| < 1$, the process has been going on for a long time, and $u_{it} \sim iid N(0, \sigma_u^2)$ the likelihood function is:

$$(2\pi)^{-nT/2} |\Omega|^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \Delta \mathbf{u}_i' \Omega^{-1} \Delta \mathbf{u}_i \right\}, \quad (3)$$

where $\Delta \mathbf{u}_i = [\Delta y_{i1}, \Delta u_{i2}, \dots, \Delta u_{iT}]'$, and the covariance matrix $\Omega = \sigma_u^2 \Omega^*$, where Ω^* is a matrix whose element (1,1) is $\frac{2}{1+\rho}$ and all other elements are -1, 0, or 2 (see Hsiao et al., 2002, formula 3.2). The corresponding Maximum likelihood estimation (MLE) is consistent as N tends to infinity, regardless of whether T is fixed or tends to infinity. The introduction of weakly exogenous variables (such as a vector x_{it}) follows a similar treatment. Moreover, if the process has started from a finite period in the past not too far back from the 0th period, and $E(\Delta y_{i1}) = b$ (where b is an unknown constant that can be estimated), the analysis is somewhat different but similar in its main spirit; see Hsiao et al. (2002) for more details.

Under the assumption of conditioning on the initial conditions, Hsiao et al. (1999) show that the Bayes estimator performs very well when T is small, which is quite important in practice. This result is certainly surprising. As Hsiao et al. (2002) mention, this could be due to the fact that the Bayes estimator is a weighted average of estimators for individual units and thus it is effectively “trimming” estimates that are unreliable in small samples. The result is certainly interesting and paves the way for consideration of more general panel data models and richer structures or sources of heterogeneity. Of course, consistent estimators may perform badly in finite samples, so consistency cannot be the sole criterion for selecting an estimator.

The reader may be surprised to see that a normal likelihood/posterior is used to provide statistical inferences without allowing for “endogeneity inherently present” in DPM. However, the endogeneity arises when the LSDV estimator is used as sample averages of the errors and the lagged dependent variables are, naturally, correlated. Moreover, for “fixed T , large N ” consistent estimators such as “first-differenced GMM” may be problematic (e.g. Bond et al., 2001) as it is not always the desired kind of asymptotic in empirical research. As a matter of fact, a fairly common assumption in dynamic panel models is that errors u_{it} , any potential regressors, and the lagged dependent variable are uncorrelated. Problems arise when the individual effects need to be wiped out by using certain transformations.

That is, problems arise when we have to face the incidental parameters problem (i.e. number of individual effects increases with n). Moral-Benito (2013) has shown that we have, essentially, a normal likelihood whose maximization is asymptotically “equivalent to the class of first-differenced GMM estimators discussed in Arellano and Bond (1991) augmented with moments resulting from lack of autocorrelation in the errors”. Additionally, provided that we make a mean-stationarity assumption, i.e. first differences of the dependent variable and regressors are orthogonal to the errors, we can exploit the resulting moment conditions as discussed in Arellano and Bover (1995). Additionally, Hsiao et al. (1999) showed that the Bayesian estimator and the familiar mean-group estimator are equivalent and consistent-asymptotically-normal estimators of the average coefficient as long as $n \rightarrow \infty, T \rightarrow \infty$ and $\sqrt{n}/T \rightarrow 0$. This implies that both n and T should be large but n should grow faster than

T so this result is more relevant in applications with large n but small T which is, admittedly, a common situation.

4. New Dynamic Models

Given the power and flexibility of the Bayesian approach, we introduce here seven new dynamic models, all developed in a Bayesian framework. Our models are based on the idea of further exhausting heterogeneity by introducing dynamic slope coefficient models. One common feature of our models is that they all allow for heterogeneity in slope parameters.¹

The base model (M1) is a model where slope coefficients are common for all cross-sectional units, and only intercepts are allowed to be different to capture heterogeneity. This can be expressed as follows:

$$y_{it} = \alpha_i + \underset{(1 \times k)}{x'_{it}} \underset{(k \times 1)}{\beta} + \rho y_{i,t-1} + u_{it}, i = 1, \dots, n, t = 1, \dots, T \quad (3)$$

where $\alpha_i \sim iidN(\bar{\alpha}, \sigma_\alpha^2)$ represents the random effects, $u_{it} \sim N(0, \sigma_u^2)$ is the error term.

The initial conditions can be expressed as: $y_{i0} \sim N(\bar{y}_{i0}, \sigma_{y_{i0}}^2)$. This is probably the simplest model that accounts for heterogeneity in the context of panel data, when $\rho=0$. Despite this fact, it is really the dynamic panel data model popularized by the Arellano and Bover and Arellano and Bond estimators. In a Bayesian context, with fixed data, estimation and inference in the model is straightforward, but this is not so in a frequentist context because of the incidental parameters problem.

Before proceeding, we need to mention that M1 is already an advanced econometric tool unless $\rho=0$, in which case we have a static panel data model which has been the main workhorse of empirical analysis for many decades. Assuming that $\rho=0$, so that we do not have dynamics, a lot of attention has been paid to whether we have fixed or random effects (which can be tested using Hausman's test). This motivates us to introduce entirely new classes of general models for panel data, which we analyze below.

4.1. Model 2 (M2): Dynamic random coefficients I

¹ This is in line with Hsiao et al. (1999). In all subsequent discussion, initial conditions are treated as unknown parameters with a $N(0, 10^2)$ prior.

Building on M1, our second model (M2) can be expressed as:

$$y_{it} = \alpha_i + \underset{(1 \times k)}{x'_{it}} \underset{(k \times 1)}{\beta_t} + \rho_t y_{i,t-1} + u_{it}, i = 1, \dots, n, t = 1, \dots, T, \quad (4)$$

where $\alpha_i \sim iidN(\bar{\alpha}, \sigma_\alpha^2)$ represents the random effects, $u_{it} \sim N(0, \sigma_u^2)$ is the error term.

The initial conditions can be expressed as: $y_{i0} \sim N(\bar{y}_{i0}, \sigma_{y_{i0}}^2)$, and we make the following assumptions for the coefficients:

$$\begin{aligned} \gamma_t &= [\beta'_t, \rho_t]', \\ \gamma_t &= \gamma_{t-1} + v_t, v_t \sim N(0, \Sigma), \\ \gamma_0 &\sim N(\bar{\gamma}_0, \lambda \Sigma), \\ \lambda &\sim Exp(\lambda_0), \lambda_0 = 1. \end{aligned} \quad (3)$$

The model is still a dynamic panel data model with intercepts that vary across individuals and slope coefficients that are allowed to be time-varying, albeit they are the same across individual units. Notably, the coefficient of lagged dependent variables are time-varying as well, which is a novelty relative to the standard Arellano and Bover and Arellano and Bond specifications. This model extends M1 in two dimensions: First, it allows for dynamics by dependence of the lagged dependent variable, and second, it allows the slope coefficients to be time-varying. Intercepts are still allowed to be different to capture heterogeneity. The time variation of the coefficients is assumed to be “smooth” in the sense that it is given by a random walk formulation.

4.2. Model 3 (M3): Dynamic random coefficients II

This model generalizes M2 in the sense that slopes and the coefficient of the lagged dependent variable are allowed to depend on both time and cross-sectional units. M3 allows for both time variation across cross-sectional units and dynamics in the slope coefficients. Again, as in M2, notably, the coefficient of lagged dependent variables is time-varying as well, which is a novelty relative to the standard Arellano and Bover and Arellano and Bond specifications. The main novelty of M3 is that it is a quite general specification for dynamic panel data in which slope coefficients are time-varying (in an autoregressive way) while they are simultaneously allowed to be different across cross-sectional units. This is clearly far more general than any specification proposed so far in the literature for handling dynamic panel data.

M3 can be expressed as follows:

$$y_{it} = \alpha_i + \underset{(1 \times k)}{\tilde{x}'_{it}} \underset{(\tilde{k} \times 1)}{\beta_{it}} + \rho_{it} y_{i,t-1} + u_{it}, i = 1, \dots, n, t = 1, \dots, T$$

or

$$y_{it} = \underset{(1 \times k)}{x'_{it}} \underset{(k \times 1)}{\beta_{it}} + \rho_{it} y_{i,t-1} + u_{it}, i = 1, \dots, n, t = 1, \dots, T,$$
(4)

where \tilde{k} is the dimensionality of the regressors in \tilde{x}_{it} , and $k = \tilde{k} + 1$. Here we assume that x_{it} contains a column of ones so that the intercept is unit-specific as well as time-varying. We rewrite the model in compact form as follows:

$$y_{it} = \underset{(1 \times k)}{z'_{it}} \underset{(k \times 1)}{\gamma_{it}} + u_{it}, i = 1, \dots, n, t = 1, \dots, T,$$
(5)

where $k = k' + 1$, and $u_{it} \sim N(0, \sigma_u^2)$ is the error term. The initial conditions can be expressed as: $y_{i0} \sim N(\bar{y}_{i0}, \sigma_{y_{i0}}^2)$, and we make the following assumptions for the coefficients:

$$\begin{aligned} \gamma_{it} &= [\beta'_{it}, \rho_{it}]', \\ \gamma_{it} &= \gamma_{i,t-1} + v_{it}, v_{it} \sim N(0, \Sigma), \\ \gamma_{i0} &\sim N(\bar{\gamma}_{i0}, \lambda_i \Sigma), \\ \lambda_i^{-1} &\sim \text{Exp}(\lambda_0), \lambda_0 = 1. \end{aligned}$$
(6)

4.4. Model 4 (M4): Mundlak device

Our fourth model takes a different approach by using the Mundlak device. The Mundlak device is used to model individual effects differently. These effects are related to average values of the regressors (so the device allows for fixed effects), but at the same time an error term is utilized so that random effects are introduced as well. Clearly, the Mundlak device is more general than either fixed or random effects and allows considerable flexibility in modeling. Since the individual effects are related to average number of regressors, the problem reduces to estimating a fixed number of parameters which do not increase with n or T so the incidental parameters problem does not arise. To illustrate, the Mundlak device generalizes fixed and random effects (for the intercept) as follows:

$$y_{it} = \alpha_i + \underset{(1 \times k)}{x'_{it}} \underset{(k \times 1)}{\beta} + \rho y_{i,t-1} + u_{it}, i = 1, \dots, n, t = 1, \dots, T$$
(7)

where $\alpha_i = \bar{x}_i' \delta + \xi_i, \xi_i \sim iidN(0, \sigma_\xi^2)$ represents the random effects, $\bar{x}_i = T^{-1} \sum_{t=1}^T x_{it}$, and $u_{it} \sim N(0, \sigma_u^2)$ is the error term, and $y_{i0} \sim N(\bar{y}_{i0}, \sigma_{y_{i0}}^2)$ are the initial conditions.

When $\sigma_\xi = 0$ we have the fixed effects formulation. When $\sigma_\xi > 0$, we have the random effects formulation augmented by an assumption of dependence of intercepts on average values of the regressors.

We can write the model as follows:

$$y_{it} = z_{it}' \gamma + (u_{it} + \xi_i), i = 1, \dots, n, t = 1, \dots, T, \quad (8)$$

where $z_{it} = [\bar{x}_i', x_{it}', y_{i,t-1}]'$. The econometric implication is that the error terms are correlated for the same unit due to the presence of ξ_i . If we stack the observations, we can write the model as

$$y_i = Z_i \gamma + v_i, \quad (9)$$

where $\text{cov}(v_i) = \sigma_u^2 I_T + \sigma_\xi^2 J_T$, and J_T denotes a matrix whose elements are all equal to one.

4.5. Model 5 (M5): Mundlak device with random coefficients I

In this specification, we augment the standard Mundlak device (M4) with the assumption of random coefficients for the slopes:

$$y_{it} = \alpha_i + \underset{(1 \times k)}{x_{it}'} \underset{(k \times 1)}{\beta_i} + \rho_i y_{i,t-1} + u_{it}, i = 1, \dots, n, t = 1, \dots, T, \quad (10)$$

where $\alpha_i = \bar{x}_i' \delta + \xi_i, \xi_i \sim iidN(0, \sigma_\xi^2)$ are the random effects, $\bar{x}_i = T^{-1} \sum_{t=1}^T x_{it}$, and $u_{it} \sim N(0, \sigma_u^2)$ is the random error term. The initial conditions can be expressed as: $y_{i0} \sim N(\bar{y}_{i0}, \sigma_{y_{i0}}^2)$, and we make the following assumptions for the coefficients:

$$\gamma_i = [\beta_i', \rho_i']' \sim N(\bar{\gamma}, \Omega), i = 1, \dots, n. \quad (11)$$

4.6. Model 6 (M6): Mundlak device with random coefficients II

Model M6 is more general than M5 and closer to the spirit of M4, in the sense that variation of slopes depends on average values of the regressors as well. This model allows for heterogeneity by modeling the slopes as functions of the average values of the regressors, instead of focusing attention exclusively on individual effects. Indeed, it can be argued that heterogeneity does not exhaust itself in modeling constant terms, as slopes can be different for each individual.

$$y_{it} = \alpha_i + \underset{(1 \times k)}{x'_{it}} \underset{(k \times 1)}{\beta_i} + \rho_i y_{i,t-1} + u_{it}, i = 1, \dots, n, t = 1, \dots, T \quad (12)$$

where $\alpha_i = \bar{x}'_i \delta + \xi_i$, $\xi_i \sim iidN(0, \sigma_\xi^2)$ are the random effects, $\bar{x}_i = T^{-1} \sum_{t=1}^T x_{it}$, and $u_{it} \sim N(0, \sigma_u^2)$ is the random error term. The advantage of this approach is that the incidental parameters problem is solved by reducing it to estimation of k parameters δ in \bar{x}' . Therefore, we extend the Mundlak device to modeling the slopes; not only the firm effects.

The initial conditions can be expressed as: $y_{i0} \sim N(\bar{y}_{i0}, \sigma_{y_{i0}}^2)$, and we make the following assumptions for the coefficients:

$$\underset{(M \times 1)}{\gamma_i} = [\alpha_i, \beta'_i, \rho_i]', \gamma_i = \begin{bmatrix} \bar{x}'_i & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \bar{x}'_i \end{bmatrix} \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_M \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_M \end{bmatrix} = (I_M \times x_i)' \underset{(Mk \times 1)}{\delta} + \underset{(Mk \times 1)}{\varepsilon} \quad (13)$$

where $\bar{X}_i = (I_M \times x_i)$ and $\varepsilon_i \sim N(0, \Omega)$, $i = 1, \dots, n$

4.7. Model 7 (M7): A general panel data model

Model M7 is general in the sense that we allow for cross-sectional and time variation of slopes:

$$y_{it} = \alpha_{it} + \underset{(1 \times k)}{x'_{it}} \underset{(k \times 1)}{\beta_{it}} + \rho_{it} y_{i,t-1} + u_{it}, i = 1, \dots, n, t = 1, \dots, T \quad (14)$$

where $\bar{x}_i = T^{-1} \sum_{t=1}^T x_{it}$, and $u_{it} \sim N(0, \sigma_u^2)$ is the random error term. The initial conditions can be expressed as: $y_{i0} \sim N(\bar{y}_{i0}, \sigma_{y_{i0}}^2)$ and we make the following assumptions for the coefficients:

$$\begin{aligned}
\gamma_{it} &= [\alpha_{it}, \beta'_{it}, \rho_{it}]', \\
& \quad (k \times 1) \\
\gamma_{it} &= \bar{\gamma} + \mu_i + \lambda_t + v_{it}, \\
& \quad (k \times 1) \quad (k \times 1) \\
v_{it} &\sim N_k(0, \Sigma), \\
\mu_i &\sim iidN_M(\mu, \Omega_\mu), \\
\lambda_t &= \lambda_{t-1} + e_t, e_t \sim iidN_k(0, \Omega_\lambda), \\
\gamma_{i0} &\sim N(\bar{\gamma}_{i0}, \lambda_i \Sigma), \lambda_0 \sim iidN_k(0, \varphi_i \Omega_\lambda), \\
\lambda_i^{-1} &\sim Exp(\lambda_0), \lambda_0 = 1, \\
\varphi_i^{-1} &\sim Exp(\varphi_0), \varphi_0 = 1.
\end{aligned} \tag{15}$$

Hence all coefficients have an additive decomposition ($\gamma_{it} = \mu_i + \lambda_t + v_{it}$) into cross-sectional effects (μ_i , time effects (λ_t) and a random component (v_{it})). The time effects evolve according to a random walk and cross-sectional effects are independent across units but dependent for the different coefficients (through matrix Ω_μ).

4.8. Model 8 (M8): A Mundlak device when the random effects depend on other functions of the data

Finally, the last model generalizes the Mundlak device in that intercepts (cross-sectional effects) do not depend exclusively on average values of the regressors but other functions of the data as well (standard deviations, higher moments, etc.):

$$y_{it} = \alpha_i + x'_{it} \beta + \rho y_{i,t-1} + u_{it}, i = 1, \dots, n, t = 1, \dots, T \tag{16}$$

(1×k) (k×1)

where $\alpha_i = \bar{x}_i' \delta_1 + \text{vech}(s_i)' \delta_2 + \xi_i$, $\xi_i \sim iidN(0, \sigma_\xi^2)$, $\bar{x}_i = T^{-1} \sum_{t=1}^T x_{it}$,

$s_i = T^{-1} \sum_{t=1}^T (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'$, and $u_{it} \sim N(0, \sigma_u^2)$ is the random error term. The initial conditions can be expressed as: $y_{i0} \sim N(\bar{y}_{i0}, \sigma_{y_{i0}}^2)$, where $\text{vech}(\cdot)$ is the vector stacking the different elements of the indicated matrix.

The model in (16) introduces more generality as the simple Mundlak device may be responsible for potential misspecification. This misspecification could be important in obtaining consistent parameter estimates as n gets larger. For example, although the Mundlak device reduces the problem of incidental parameters by estimating a *fixed* number of parameters, the parameters relate to *average* values of the regressors. This is, however, only an assumption; if it is not true, misspecification will compromise the ability of the model to deliver consistent estimators. From the point of view of modeling heterogeneity, the assumption that individual effects relate to average values of the regressors may be wrong when the distribution of regressors matters for individual effects. This can be taken into account, while not compromising the reduction of the problem of incidental parameters to estimating a fixed number of parameters, by including higher moments of the distribution of regressors into the Mundlak device. For example, standard deviations or variances of the regressors can be used to ‘explain’ the individual effects, in the interest of making the approach more flexible.

5. Empirical Application

We test our models using an application on the impact of advertising on firm sales. Most previous studies in the literature (Heyse and Wei, 1985; Sonnier et al. 2011; Bruce et al. 2012; Assaf et al. 2015) have assumed a dynamic impact of advertising on sales (or other related performance metrics) due to the carry-over effects or wear out of advertising campaigns (Bass et al. 2007). In our estimation, we focused on a sample of US restaurants, covering a sample of 22 publicly traded restaurants from 2001 to 2016². The data is unbalanced and we have a total of 341 observations³. In addition to advertising, we also control for the impact of firm size and financial leverage, two variables that are commonly used in similar contexts (Luo et al., 2010; Lee et al., 2013). All data we collected from the COMPUSTAT database. Following previous research (e.g. Luo and De Jong, 2012), we measured advertising spending as the reported firm advertising expenditure. We measured firm size as the log of total assets and financial leverage as firms’ long-term book debt over total assets. We provide the descriptive statistics of all these variables in Table 1.

6. Results

As mentioned, we estimated all our models in a Bayesian framework. Appendix A provides more technical about the Bayesian procedure of the various models. For the purpose of comparison, we estimate all eight models proposed in this paper. We use the model pooling proposed by Geweke and Amisano (2011). This approach assumes that none of the competing models corresponds to the true data generating process, but instead considers a linear

² The financial data was not available in some years for all the restaurants included in the sample, resulting in the unbalanced sample.

³ See footnote 2.

prediction pool based on the predictive likelihood (log score function) from a set of competing models. To illustrate, given a set of models $\{\mathcal{M}_i\}_{i=1}^M$ and a set of predictive densities $\{p(y_t|y_1, \dots, y_{t-1}, \mathcal{M}_i)\}_{i=1}^M$, we consider the following form of combined predictive densities:

$$\sum_{i=1}^M w_i p(y_t|y_1, \dots, y_{t-1}, \mathcal{M}_i), \text{ where } \sum_{i=1}^M w_i = 1, w_i \geq 0, i = 1, \dots, M. \quad (17)$$

The optimal weight vector \mathbf{w}^* is chosen to maximise the log pooled predictive score function; that is,

$$\operatorname{argmax}_{w_i, i=1, \dots, M} \sum_{t=\tau_1}^{\tau_2} \log \left(\sum_{i=1}^M w_i p(y_t|y_1, \dots, y_{t-1}, \mathcal{M}_i) \right), \quad (18)$$

where the predictive density is evaluated at the realized value y_t . Conditional on the data up to time $t - 1$, i.e., y_1, \dots, y_{t-1} , we obtain a large number of posterior draws for the parameters (by applying the MCMC algorithm for a large number of iterations), which are then used to evaluate the predictive likelihood $p(y_t = y_t^o | y_1, \dots, y_{t-1}, \mathcal{M}_i)$. Based on the entire history of predictive likelihood values we can estimate the weights in expression (18). For optimization, we have used a standard nonlinear solver (Nash, 1984). Given the data Y we define the marginal likelihood or evidence of a model as:

$$M_i(Y) = \int L_i(\theta_i; Y) p_i(\theta_i) d\theta_i, \quad (19)$$

where $L_i(\theta_i; Y)$, $p_i(\theta_i)$ denote the likelihood and prior, respectively.

Using (19) we can also construct the posterior model probabilities (PMP), which given (in-sample) marginal likelihoods $M_i(Y)$, $i = 1, \dots, M$, the PMP for model i can be defined as:

$$PMP_i = \frac{M_i(Y)}{\sum_{j=1}^M M_j(Y)}, i = 1, \dots, M. \quad (20)$$

Table 2 reports the model comparison's results. Based on the Bayes factor, the model pooling of Geweke and Amisano, and PMP, M8 outperforms all other models in our application. For example, in comparison to the base model M1, M8 has a Bayes factor of 103.21, indicating a considerably better performance. According to model pooling, M8 has the highest PMP, clearly indicating that it performs best, followed then by other models with $PMP > 0$.

Further evidence on the performance of the various models is reported in Table 3 where we provide the out of sample forecasting performance of each model. The magnitude of forecasting mistakes is measured with root mean squared forecast error (*RMSFE*). Forecasting is performed for one and two years ahead. Specifically, for each hotel, we leave out the last one or two years, estimate the models and predict the left-out observations using the posterior mean point forecast and compute the *RMSFEs* as usual. We also tested forecasting in two other ways. First, we use the Geweke-Amisano out-of-sample posterior model probabilities to predict observations that we left out. Second, we use the in-sample posterior model probabilities to perform Bayesian model averaging and, in turn, predict out of sample. Clearly, the models are performing well across both approaches, but the first approach clearly performs much better as the *RMSFEs* are much smaller. In addition, it is clear that M8 has the lowest *RMSFE* in our present application, indicating best forecasting performance.

Finally, we report (in Table 4) the posterior mean and standard deviations of all variables in our application. These include advertising, size, leverage, as well as the lag of sales. Figures 1-4 also report the posterior densities of these variables. It is clear that all models indicate a significant and positive impact of advertising on sales, as is expected from the literature (e.g. Darrat et al. 2015; McAlister et al. 2016). The lag of sales also seems significant across most models, providing more support for the use of dynamic formulation. As is clear from the densities, there are notable deviations from normality, so asymptotic theory may not be valid in our context. In Figure 5, we report the same densities using Bayesian model averaging (BMA), where we average over all models used in the paper. This way, we account for any uncertainty (when assessing parameters and predictions) that can result from the model selection process. In general, our results are consistent with those reported in Figures 1-4. For instance, we can see from Figure 5 the effect of advertising on sales is positive and close to 0.23 ranging from 0.17 to 0.3.

7. Conclusions

In this paper, we introduced several innovative dynamic panel data models that allow for variations in slope coefficients both across time and cross-sectional units. To our knowledge, these models have not been explored in previous studies. As mentioned, our models better capture heterogeneity, as most current static or dynamic panel data models are based on the assumption that slope coefficients are common across units and heterogeneity is modeled using fixed or random effects for the intercept. However, heterogeneity is not fully captured through assuming that intercepts are unit-specific or by using fixed or random effect estimators. In our models, we allow the slope coefficients to vary both across time and cross-sectional units, and in a dynamic fashion (i.e. dynamic slope coefficients). Importantly, we also introduced for the first time, flexible models based on the Mundlak device; which is more general than either fixed or random effects and does not suffer from the incidental parameters problem.

We applied these new models in an interesting application showing the effects of advertising on sales. We compared their performance using model pooling and posterior model probabilities. We also assessed their out of sample forecasting performance. For the present application it was clear that M8 outperformed all other models, which is not surprising given that M8 is based on the more flexible Mundlak device and is more general in the sense that it is likely to better address any misspecification issues created by the simple Mundlak device. Regardless, these models are open for testing, and future studies are encouraged to at least consider some of them for comparison purposes. Simply, relying on the Arellano-Bond (AB) model, as is common in the tourism literature, may run the risk of misspecification in some applications, as the AB model (in contrast to the models proposed in this study) is based on fixed slope coefficients. Hence, it does not fully exhaust heterogeneity.

Finally, we encourage more use of the Bayesian approach for the estimation of dynamic panel data models in tourism research. As mentioned, the Bayesian approach has particular advantages in the context of dynamic models, as the use of lagged dependent variables does not introduce any additional problems with Bayesian estimation. In addition, the use of the GMM estimator is problematic when the Arellano-Bond instruments are weak. In such a case, the sampling behavior of GMM can lead to unreliable estimates.

Table 1. Descriptive Statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
Sales	341	3251.621	5445.581	52.346	28105.7
Advertising	341	101.722	181.564	0.1	808.4
Firm Size	341	2.971	0.588	1.477	4.579
Financial Leverage	341	0.670	0.600	0.144	4.071

Table 2. Model Comparison

	M1	M2	M3	M4	M5	M6	M7	M8
Bayes factor	1.000	7.44	11.28	15.62	11.03	27.36	81.16	103.21
Geweke-Amisano	0.000	0.000	0.000	0.000	0.000	0.038	0.092	0.870
PMP	0.0039	0.0288	0.0437	0.0605	0.0427	0.106	0.3145	0.3999

Notes: Geweke-Amisano provides the Geweke-Amisano posterior model probabilities in a posterior predictive pool using a hold-out sample of five observations for each firm. PMP denotes posterior model probabilities using the Bayes factor.

Table 3. Out of sample forecasting

	RMSFE (%), 1 year ahead	RMSFE (%), 2 years ahead
M1	34.41	40.01
M2	22.32	32.60
M3	17.51	22.35
M4	12.35	16.71
M5	7.31	9.44
M6	2.45	3.08
M7	1.93	2.79
M8	1.81	2.44
Geweke-Amisano pool	1.32	2.05
Averaging based on in-sample BF	2.33	3.71

Table 4. Posterior Mean and Standard Deviation

	M1	M2	M3	M4	M5	M6	M7	M8
advertising	0.209* (0.017)	0.313* (0.032)	0.144* (0.017)	0.120* (0.009)	0.233* (0.024)	0.187* (0.017)	0.302* (0.015)	0.221* (0.015)
size	1.762* (0.07)	1.351* (0.025)	0.856* (0.022)	0.921* (0.019)	1.150* (0.022)	0.933* (0.019)	1.230* (0.026)	1.190* (0.035)
leverage	-0.019 (0.032)	-0.044 (0.030)	-0.035* (0.012)	-0.022 (0.030)	-0.015* (0.007)	-0.022* (0.006)	-0.017* (0.002)	-0.031 (0.017)
lagged sales	0.315* (0.021)	0.521* (0.018)	0.552* (0.013)	0.613* (0.022)	0.587* (0.019)	0.616* (0.022)	0.510* (0.035)	0.481* (0.029)

Note: Numbers in parentheses are the posterior standard deviations. * indicates significance at the 5% level.

FIGURE 1. Marginal Posterior Densities of Advertising across All Models

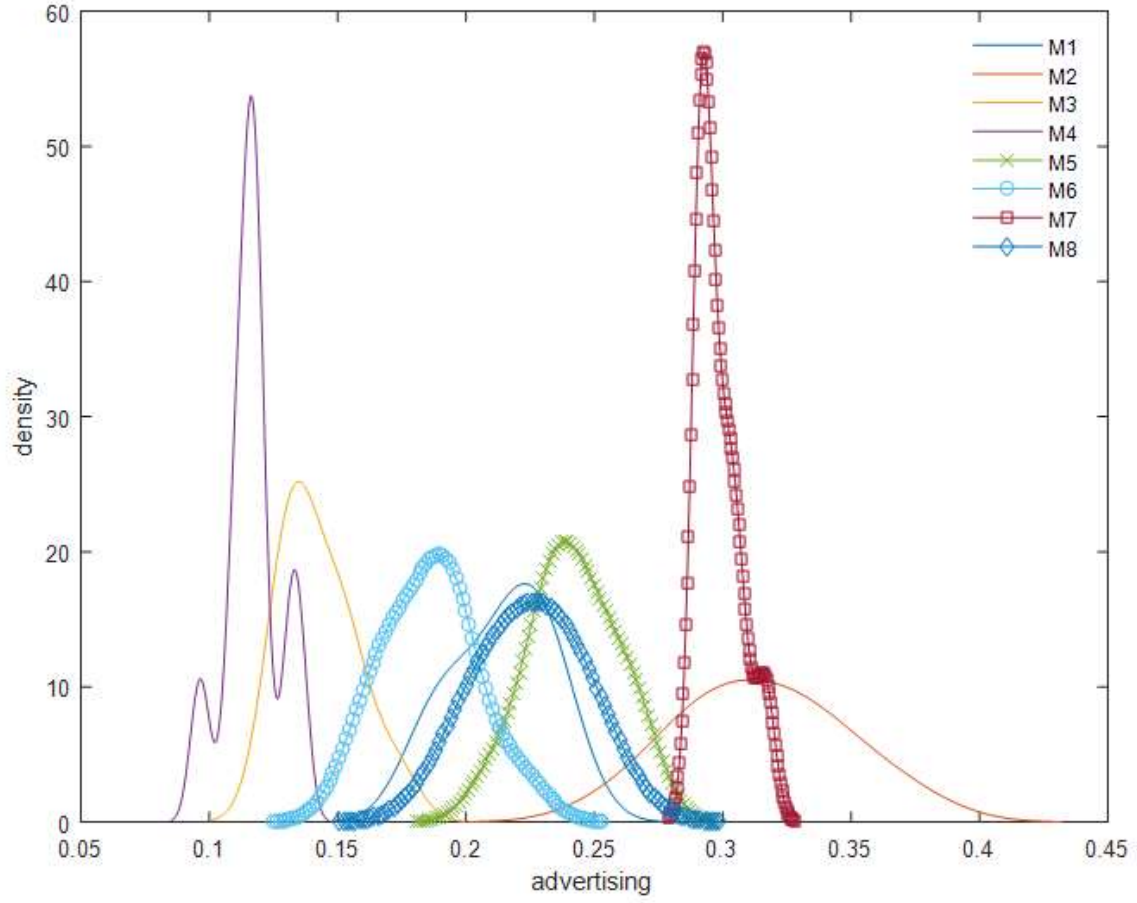


FIGURE 2. Marginal Posterior Densities of Firm Size across All Models

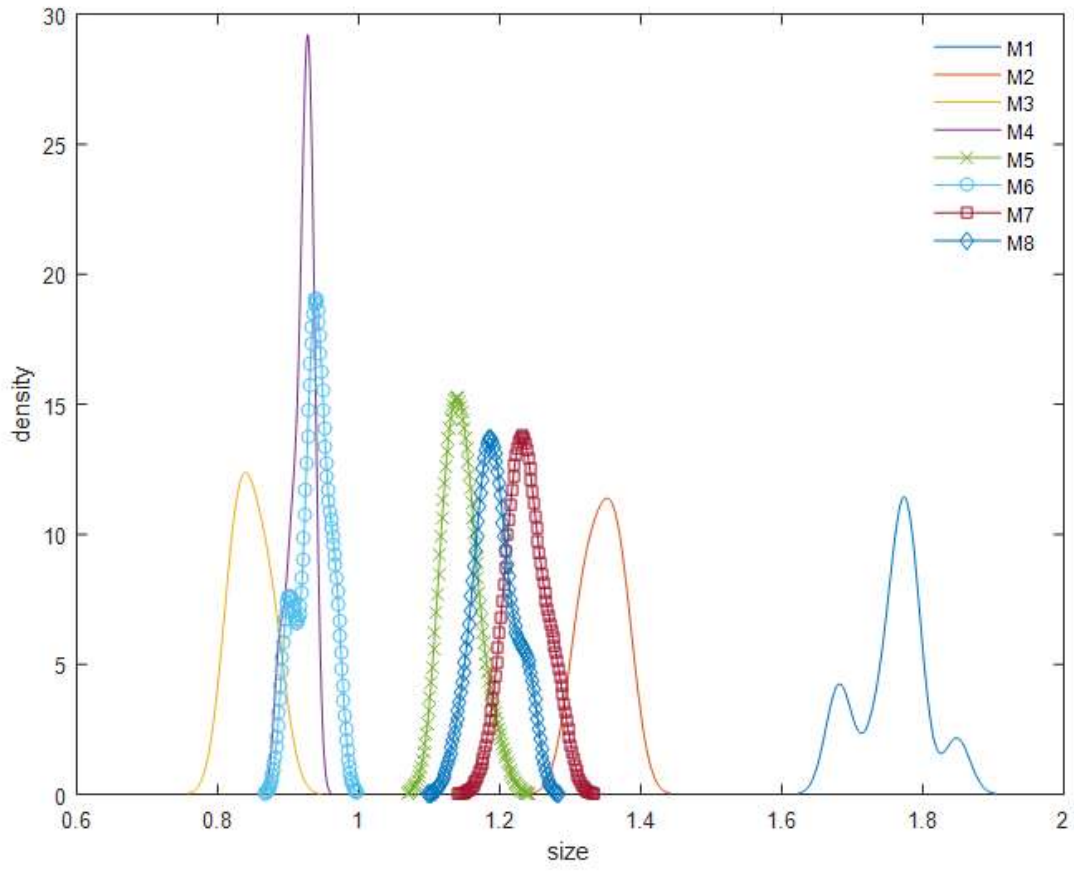


FIGURE 3. Marginal Posterior Densities of Leverage across All Models

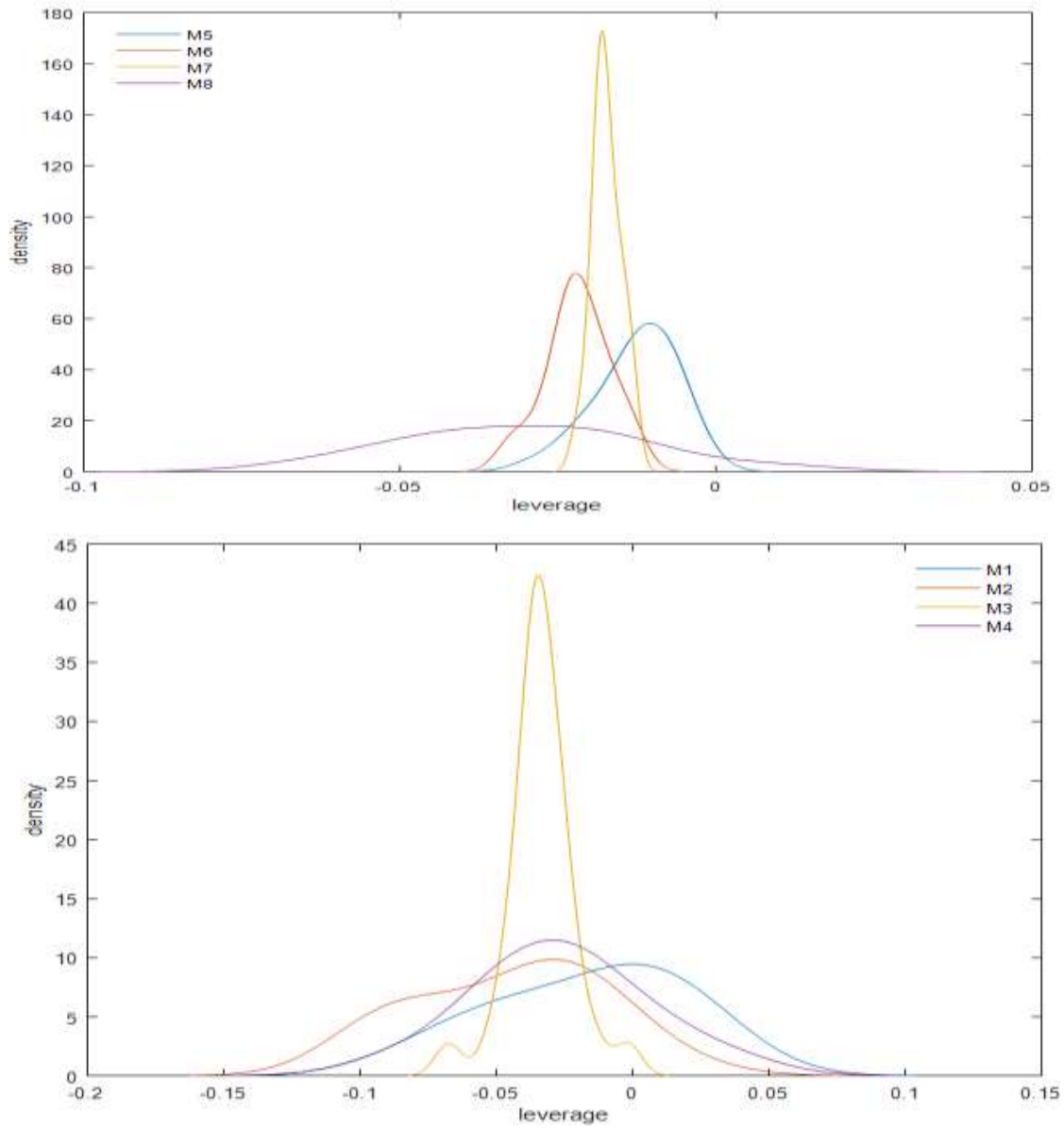


FIGURE 4. Marginal Posterior Densities of Lagged Sales across All Models

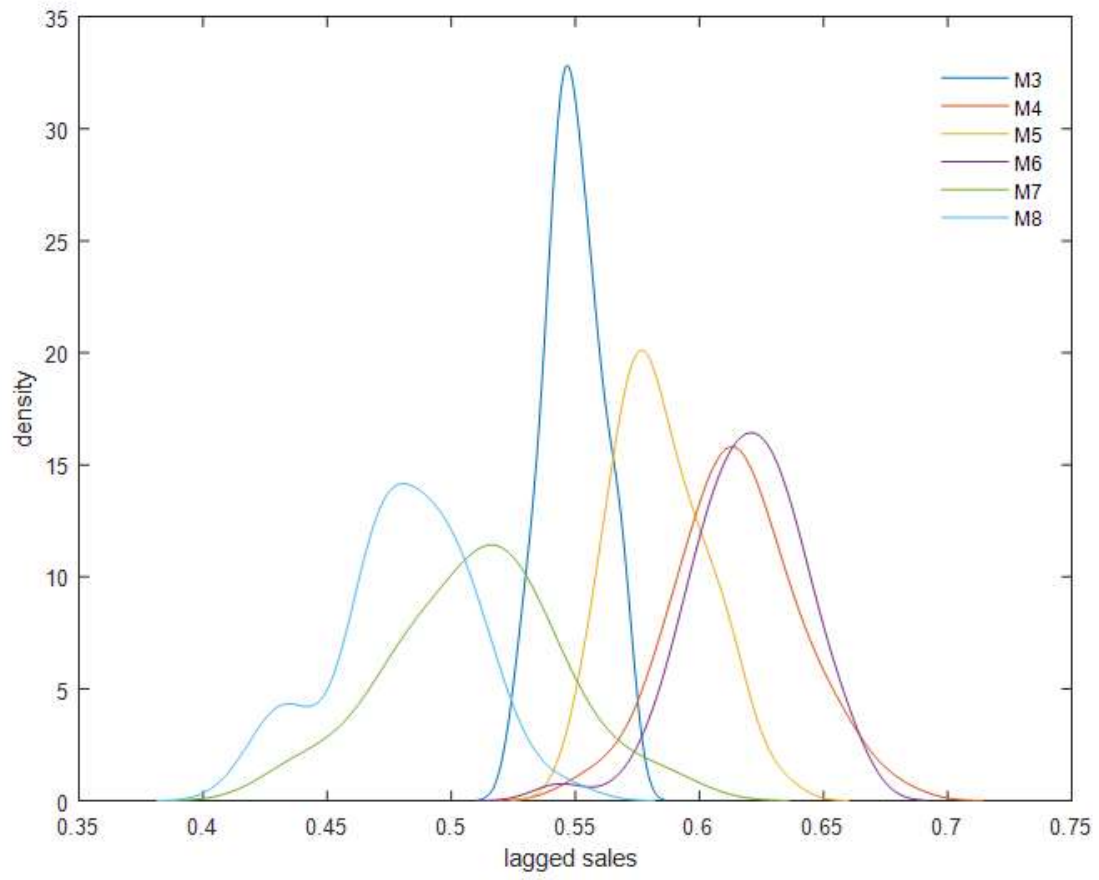
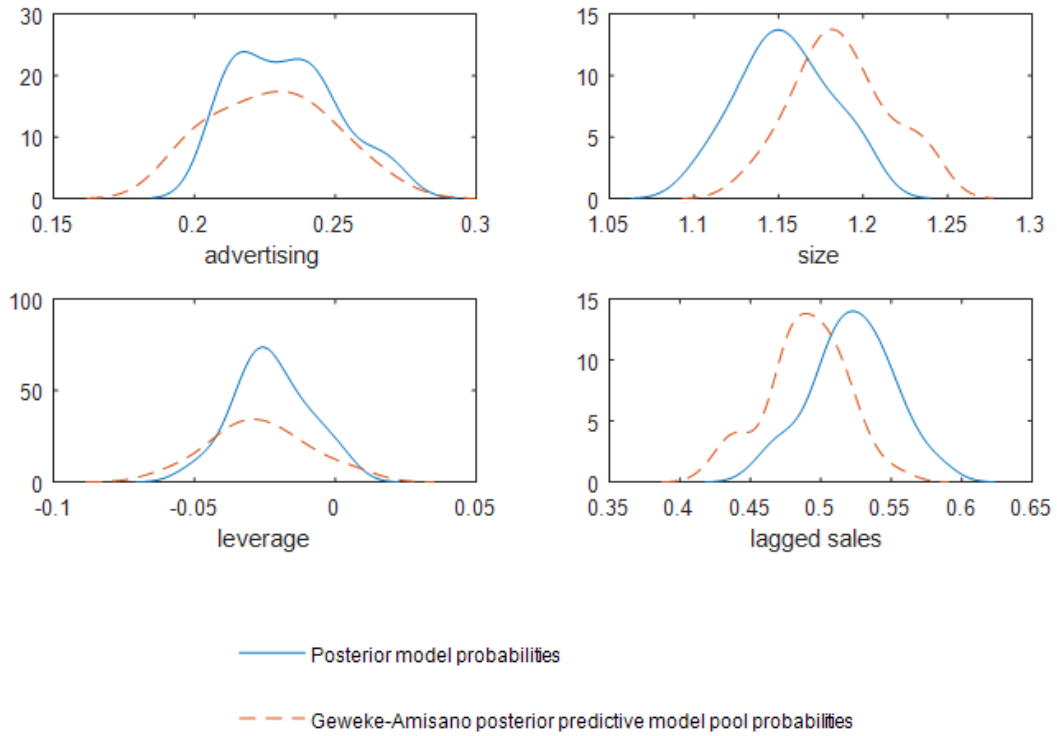


FIGURE 5. Model-averaged Marginal Posterior Densities



8. References

- Ahn, S. C., & Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. *Journal of econometrics*, 68(1), 5-27.
- Anderson, T. W., & Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American statistical Association*, 76(375), 598-606.
- Anderson, T. W., & Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of econometrics*, 18(1), 47-82.
- Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The review of economic studies*, 58(2), 277-297.
- Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of econometrics*, 68(1), 29-51.
- Assaf, A. G., Josiassen, A., Mattila, A. S., & Cvelbar, L. K. (2015). Does advertising spending improve sales performance? *International Journal of Hospitality Management*, 48, 161-166.
- Assaf, A. G., & Tsionas, M. G. (2019). A review of research into performance modeling in tourism research-Launching the Annals of Tourism Research curated collection on performance modeling in tourism research. *Annals of Tourism Research*, 76, 266-277.
- Bilen, M., Yilanci, V., & Eryüzlü, H. (2017). Tourism development and economic growth: a panel Granger causality analysis in the frequency domain. *Current Issues in Tourism*, 20(1), 27-32.
- Bass, F. M., Bruce, N., Majumdar, S., & Murthi, B. P. S. (2007). Wearout effects of different advertising themes: A dynamic Bayesian model of the advertising-sales relationship. *Marketing Science*, 26(2), 179-195.
- Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1), 115-143.
- Brida, J. G., & Risso, W. A. (2009). A dynamic panel data study of the German demand for tourism in South Tyrol. *Tourism and Hospitality Research*, 9(4), 305-313.
- Bruce, N. I., Foutz, N. Z., & Kolsarici, C. (2012). Dynamic effectiveness of advertising and word of mouth in sequential distribution of new products. *Journal of Marketing Research*, 49(4), 469-486.
- Darrat, M. A., Wilcox, G. B., Funches, V., & Darrat, M. A. (2016). Toward an understanding of causality between advertising and sales: New evidence from a multivariate cointegrated system. *Journal of Advertising*, 45(1), 62-71.

- Du, D., Lew, A. A., & Ng, P. T. (2016). Tourism and economic growth. *Journal of Travel Research*, 55(4), 454-464.
- Falk, M. (2010). A dynamic panel data analysis of snow depth and winter tourism. *Tourism Management*, 31(6), 912-924.
- Fayissa, B., Nsiah, C., & Tadasse, B. (2008). Impact of tourism on economic growth and development in Africa. *Tourism Economics*, 14(4), 807-818.
- Garín-Mun, T. (2006). Inbound international tourism to Canary Islands: a dynamic panel data model. *Tourism management*, 27(2), 281-291.
- Geweke, J., & Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164, 130–141.
- Heyse, J. F., & Wei, W. W. (1985). Modelling the advertising-sales relationship through use of multiple time series techniques. *Journal of Forecasting*, 4(2), 165-181.
- Hsiao, C., Pesaran, M.H. and Tahmiscioglu, A.K. (1999) "Bayes Estimation of Short-Run Coefficients in Dynamic Panel Data Models" in (eds.) C. Hsiao, K. Lahiri, L-F Lee and M.H. Pesaran, *Analysis of Panels and Limited Dependent Variables: A Volume in Honour of G S Maddala*, Cambridge University Press, Cambridge, 1999, chapter 11, pp.268-296.
- Hsiao, C., Pesaran, M. H., & Tahmiscioglu, A. K. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of econometrics*, 109(1), 107-150.
- Kim, J., Jun, J., & Tang, L. R. (2019). How well does advertising work on restaurant performance? A dynamic and quadratic approach. *International Journal of Hospitality Management*, 81, 11-20.
- Lee, S., Seo, K., & Sharma, A. (2013). Corporate social responsibility and firm performance in the airline industry: The moderating role of oil prices. *Tourism Management*, 38, 20-30.
- Li, H., Goh, C., Zhang Qiu, H., & Meng, F. (2015). Effect of tourism on balanced regional development: A dynamic panel data analysis in Coastal and Inland China. *Asia Pacific Journal of Tourism Research*, 20(6), 694-713.
- Li, H., Song, H., & Li, L. (2017). A dynamic panel data analysis of climate and tourism demand: Additional evidence. *Journal of Travel Research*, 56(2), 158-171.
- Luo, X., & de Jong, P. J. (2012). Does advertising spending really work? The intermediate role of analysts in the impact of advertising on firm value. *Journal of the Academy of Marketing Science*, 40(4), 605-624.
- Luo, X., Homburg, C., & Wieseke, J. (2010). Customer satisfaction, analyst stock recommendations, and firm value. *Journal of Marketing Research*, 47(6), 1041-1058.

- Maloney, W. F., & Montes Rojas*, G. V. (2005). How elastic are sea, sand and sun? Dynamic panel estimates of the demand for tourism. *Applied Economics Letters*, 12(5), 277-280.
- Massidda, C., & Etzo, I. (2012). The determinants of Italian domestic tourism: A panel data analysis. *Tourism Management*, 33(3), 603-610.
- McAlister, L., Srinivasan, R., Jindal, N., & Cannella, A. A. (2016). Advertising effectiveness: the moderating effect of firm strategy. *Journal of Marketing Research*, 53(2), 207-224.
- Moral-Benito, E. (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. *Journal of Business & Economic Statistics*, 31(4), 451-472.
- Nash, S.G. (1984). Newton-type minimization via the Lanczos method. *SIAM Journal on Numerical Analysis*, 21, 770-788.
- Naudé, W. A., & Saayman, A. (2005). Determinants of tourist arrivals in Africa: a panel data regression analysis. *Tourism Economics*, 11(3), 365-391.
- Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1), 1-32.
- Paramati, S. R., Alam, M. S., & Chen, C. F. (2017). The effects of tourism on economic growth and CO2 emissions: a comparison between developed and developing economies. *Journal of Travel Research*, 56(6), 712-724.
- Rey, B., Myro, R. L., & Galera, A. (2011). Effect of low-cost airlines on tourism in Spain. A dynamic panel data model. *Journal of Air Transport Management*, 17(3), 163-167.
- Saha, S., Su, J. J., & Campbell, N. (2017). Does political and economic freedom matter for inbound tourism? A cross-national panel data estimation. *Journal of Travel research*, 56(2), 221-234.
- Santana-Gallego, M., Ledesma-Rodríguez, F., & Pérez-Rodríguez, J. V. (2011). Tourism and trade in OECD countries. A dynamic heterogeneous panel data analysis. *Empirical Economics*, 41(2), 533.
- Seetanah, B. (2011). Assessing the dynamic economic impact of tourism for island economies. *Annals of Tourism Research*, 38(1), 291-308.
- Seetaram, N. (2010). Use of dynamic panel cointegration approach to model international arrivals to Australia. *Journal of Travel Research*, 49(4), 414-422.
- Seetaram, N., & Petit, S. (2012). Panel data analysis. *Handbook of research methods in tourism: Quantitative and qualitative approaches*, 127-144.

Sequeira, T. N., & Maçãs Nunes, P. (2008). Does tourism influence economic growth? A dynamic panel data approach. *Applied Economics*, 40(18), 2431-2441.

Sonnier, G. P., McAlister, L., & Rutz, O. J. (2011). A dynamic model of the effect of online communications on firm sales. *Marketing Science*, 30(4), 702-716.

Woo, L., Assaf, A. G., Josiassen, A., & Kock, F. (2019). Internationalization and hotel performance: Agglomeration-related moderators. *International Journal of Hospitality Management*, 82, 48-58.

Wu, T. P., & Wu, H. C. (2019). Tourism and economic growth in Asia: A bootstrap multivariate panel Granger causality. *International Journal of Tourism Research*, 21(1), 87-96.

Yang, Y. (2012). Agglomeration density and tourism development in China: An empirical research based on dynamic panel data model. *Tourism Management*, 33(6), 1347-1359.

Zhang, L., & Gao, J. (2016). Exploring the effects of international tourism on China's economic growth, energy consumption and environmental pollution: Evidence from a regional panel analysis. *Renewable and Sustainable Energy Reviews*, 53, 225-234.

Appendix A: Bayesian Technical details

This appendix provides more technical details about Bayesian estimation. We focus on M3 which is a generalization of M1 and M2, M6, which is a generalization of M4 and M5, and M7 which can be simply extended to M8.

Model 3 (M3)

Considering M3 again as defined in (5):

$$y_{it} = \underset{(1 \times k)}{z'_{it}} \underset{(k \times 1)}{\gamma_{it}} + u_{it}, i = 1, \dots, n, t = 1, \dots, T, \quad (\text{A.1})$$

The augmented posterior of the model can be written as follows:

$$\begin{aligned} p(\sigma_u, \Sigma, \{\gamma_{it}\}, \{\lambda_i\} | y, X) &\propto \sigma_u^{-(nT+1)} \exp\left\{-\frac{1}{2\sigma_u^2} \sum_{i=1}^n \sum_{t=1}^T (y_{it} - z'_{it}\gamma_{it})^2\right\} \\ &|\Sigma|^{-(nT+k+1)/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^n \sum_{t=1}^T (\gamma_{it} - \gamma_{i,t-1})' \Sigma^{-1} (\gamma_{it} - \gamma_{i,t-1})\right\} \\ &\left(\prod_{i=1}^n |\lambda_i \Sigma|^{-1/2}\right) \exp\left\{-\frac{1}{2} \sum_{i=1}^n (\gamma_{i0} - \bar{\gamma}_{i0})' (\lambda_i \Sigma)^{-1} (\gamma_{i0} - \bar{\gamma}_{i0})\right\} \\ &\exp\left\{-\lambda_0 \sum_{i=1}^n \lambda_i^{-1}\right\} \end{aligned} \quad (\text{A.2})$$

where our prior is the standard non-informative form:

$$p(\sigma_u, \Sigma) \propto \sigma_u^{-1} |\Sigma|^{-(k+1)/2}. \quad (\text{A.3})$$

Drawing from the conditional posteriors of σ_u, Σ is straightforward since:

$$\frac{\sum_{i=1}^n \sum_{t=1}^T (y_{it} - z'_{it}\gamma_{it})^2}{\sigma_u^2} | \bullet, y, X \sim \chi^2(nT), \quad (\text{A.4})$$

and

$$p(\Sigma | \bullet) \propto |\Sigma|^{-(nT+k+2)/2} \exp\left\{-\frac{1}{2} \text{tr} \Sigma^{-1} A\right\}, \quad (\text{A.5})$$

where $A = \sum_{i=1}^n \sum_{t=1}^T (\gamma_{it} - \gamma_{i,t-1})(\gamma_{it} - \gamma_{i,t-1})' + \sum_{i=1}^n \lambda_i (\gamma_{i0} - \bar{\gamma}_{i0})(\gamma_{i0} - \bar{\gamma}_{i0})'$. This conditional posterior is in the inverted Wishart family.

The conditional posterior of each λ_i is:

$$p(\lambda_i | \bullet) \propto \lambda_i^{-k/2} \exp(-a\lambda_i^{-1}), i = 1, \dots, n, \quad (\text{A.6})$$

where $a = (\gamma_{i0} - \bar{\gamma}_{i0})' \Sigma^{-1} (\gamma_{i0} - \bar{\gamma}_{i0}) + \lambda_0$. This is an inverted gamma distribution provided $k > 2$.

The difficult part of the MCMC sampler is to draw efficiently $\{\gamma_{it}\}$. These parameters can be updated as follows:

$$\gamma_{it} | \gamma_{i,t-1}, \gamma_{i,t+1}, \bullet, y, X \sim N_k(\hat{\gamma}_{it}, \hat{V}_{it}), i = 1, \dots, n, t = 1, \dots, T-1, \quad (\text{A.7})$$

where $\hat{\gamma}_{it} = (z_{it}' z_{it}' + 2\sigma_u^2 \Sigma^{-1})^{-1} [z_{it}' y_{it} + \sigma_u^2 \Sigma^{-1} (\gamma_{i,t-1} + \gamma_{i,t+1})]$, $t = 1, \dots, T-1$, and $\hat{V}_{it} = \sigma_u^2 (z_{it}' z_{it}' + 2\sigma_u^2 \Sigma^{-1})^{-1}$ with the proper adjustments for γ_{i0} and γ_{iT} .

Model M6 (Mundlak device)

We use (12) to write the model in the form:

$$\begin{aligned} y_{it} &= z_{it}' \beta_i + u_{it}, i = 1, \dots, n, t = 1, \dots, T, \\ \beta_i &= \bar{X}_i \delta + \varepsilon_i, i = 1, \dots, n, \end{aligned} \quad (\text{A.8})$$

If we stack observations, we have $y_i = Z_i \beta_i + u_i$, and using the Mundlak device we obtain:

$$y_i = (Z_i \bar{X}_i) \delta + (u_i + Z_i \varepsilon_i), i = 1, \dots, n. \quad (\text{A.9})$$

This is a model with fixed coefficients and the covariance of the error is

$$E \left[(u_i + Z_i \varepsilon_i)(u_i + Z_i \varepsilon_i)' \right] = \sigma_u^2 I_T + Z_i \Omega Z_i', i = 1, \dots, n, \quad (\text{A.10})$$

which shows that it is different for different units. The likelihood function of the model can be written as follows⁴:

$$L(\theta; y, X) = \prod_{i=1}^n |\sigma_u^2 I_T + Z_i \Omega Z_i'|^{-1/2} \cdot \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (y_i - (Z_i \bar{X}_i) \delta)' (\sigma_u^2 I_T + Z_i \Omega Z_i')^{-1} (y_i - (Z_i \bar{X}_i) \delta) \right\} \quad (\text{A.11})$$

⁴ For simplicity we omit terms involving the initial conditions.

where the parameter vector $\theta = [\delta', \sigma_u, \text{vech}(\Omega)']'$. If we parametrize Ω in terms of the non-zero elements of its Cholesky factorization, $\Omega = C'C$ where C is lower triangular, it is not difficult to use general-purpose optimization algorithms to find the maximum likelihood estimator.

Given a non-informative prior, $p(\theta) \propto \sigma_u^{-1} |\Omega|^{-(k+1)/2}$, where with some abuse of notation in the main text, k is the dimensionality of z_{it} , the posterior can be analyzed efficiently using the Girolami and Calderhead (2012) algorithm. The algorithm uses first and second derivative information about the log-posterior: $p(\theta | y, X) \propto L(\theta; y, X) p(\theta)$. Given a MCMC draw for θ , a draw for γ_i can be obtained as:

$$\gamma_i | \bullet, y, X \sim N(\hat{\gamma}_i, \hat{V}_i), i = 1, \dots, n, \quad (\text{A.12})$$

$$\text{where } \hat{\gamma}_i = \left(Z_i' Z_i + \sigma_u^2 \Omega^{-1} \right)^{-1} \left(Z_i' y_i + \sigma_u^2 \Omega^{-1} \bar{X}_i \delta \right), \quad \hat{V}_i = \sigma_u^2 \left(Z_i' Z_i + \sigma_u^2 \Omega^{-1} \right)^{-1}, i = 1, \dots, n.$$

Notice that, in this instance, we “concentrate out” the unit-specific parameters as we are using the likelihood $L(\theta; y, X)$. Of course, the algorithm can be used in several special cases including Model M3. For Model M8 we use exactly the same algorithm as the only difference relative to Model M6 concerns the definition of \bar{X}_i .

Model M7

Model M7 has the following structure:

$$y_{it} = z_{it}' \gamma_{it} + u_{it}, i = 1, \dots, n, t = 1, \dots, T, \quad (\text{A.13})$$

with

$$\gamma_{it} = \gamma + \mu_i + \lambda_t + v_{it} \quad (\text{A.14})$$

Conditionally on the time effects, we have:

$$y_{it} - z_{it}' \lambda_t = z_{it}' \bar{\gamma} + z_{it}' (v_{it} + \mu_i) + u_{it} \quad (\text{A.15})$$

which we can stack as follows:

$$y_i - Z_i \lambda = Z_i \gamma + \varepsilon_i, i = 1, \dots, n \quad (\text{A.16})$$

where $\varepsilon_i = Z_i (v_i + \mu_i) + u_i$ whose covariance matrix is

$$\Sigma_i = E(\varepsilon_i \varepsilon_i') = Z_i (\Sigma + \Omega_\mu) Z_i' + \sigma_u^2 I \quad (\text{A.17})$$

Therefore, we can express the posterior as

$$p(\theta | y, X) \propto \int \left[\prod_{i=1}^n |\Sigma_i|^{-1/2} \right] \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (y_i - Z_i \lambda - Z_i \gamma)' \Sigma_i^{-1} (y_i - Z_i \lambda - Z_i \gamma) \right\} \cdot |\Omega_\lambda|^{-T/2} \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (\lambda_t - \lambda_{t-1})' \Omega_\lambda^{-1} (\lambda_t - \lambda_{t-1}) \right\} d\lambda \cdot p(\theta) \quad (\text{A.18})$$

where $p(\theta)$ is a non-informative prior. Conditional on λ , we can draw γ as follows:

$$\gamma | \cdot, y, X \sim N \left((Z' \tilde{\Sigma}^{-1} Z)^{-1} Z' \tilde{\Sigma}^{-1} (Y - Z \lambda), (Z' \tilde{\Sigma}^{-1} Z)^{-1} \right), \quad (\text{A.19})$$

where $Z = \text{diag}[Z_1, \dots, Z_n]$, $Y = [Y_1', \dots, Y_n']'$ and $\tilde{\Sigma} = \text{diag}[\Sigma_1, \dots, \Sigma_n]$. The different elements of Σ, Ω_μ and σ_u^2 are updated using a Girolami and Calderhead (2012) technique (see Appendix B). The different elements of Ω_λ are updated using the Wishart conditional posterior:

$$p(\Omega_\lambda | \cdot, y, X) \propto |\Omega_\lambda|^{-(T+k+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} A_\lambda \Omega_\lambda^{-1} \right\}, \quad (\text{A.20})$$

$$\text{where } A_\lambda = \sum_{t=1}^T (\lambda_t - \lambda_{t-1}) (\lambda_t - \lambda_{t-1})'.$$

To draw from the posterior conditional of λ_t we write the model as follows:

$$y_{it} - z_{it}' (\bar{\gamma} + \mu_i) = z_{it}' \lambda_t + (z_{it}' v_{it} + u_{it}).$$

We can stack the observations as

$$\tilde{y}_t \equiv y_t - Z_t (\bar{\gamma} + \mu_i) = Z_t \lambda_t + (Z_t v_t + u_t), t = 1, \dots, T, \quad (\text{A.21})$$

where $\text{cov}(Z_t v_t + u_t) = Z_t (I_n \otimes \Sigma) Z_t' + \sigma_u^2 I_n \equiv V_t$. Since $\lambda_t | \lambda_{t-1}, \Omega_\lambda \sim N(\lambda_{t-1}, \Omega_\lambda)$ the conditional posterior of time effects is given as follows:

$$\lambda_t | \lambda_{t-1}, \lambda_{t+1}, \cdot, y, X \sim N(\hat{\lambda}_t, \hat{V}_t), \quad (\text{A.22})$$

where

$$\hat{\lambda}_t = \left[Z_t' V_t^{-1} Z_t + 2\Omega_\lambda^{-1} \right]^{-1} \left[Z_t' V_t^{-1} \tilde{y}_t + \Omega_\lambda^{-1} (\lambda_{t-1} + \lambda_{t+1}) \right], t = 1, \dots, T-1, \text{ and}$$

$$\hat{V}_t = \left[Z_t' V_t^{-1} Z_t + 2\Omega_\lambda^{-1} \right]^{-1}$$

For λ_0, λ_T we need the obvious adjustments in the conditional posterior distributions.

Appendix B: The Girolami-Calderhead Update

Suppose $L(\theta) = \log p(\theta | \mathbf{X})$ is used to denote (for simplicity) the log posterior of θ . Let us define

$$\mathbf{G}(\theta) = \text{est.cov} \frac{\partial}{\partial \theta} \log p(\mathbf{X} | \theta) \quad (\text{A.23})$$

which is the sample analogue of

$$\mathbf{G}_o(\theta) = -E_{Y|\theta} \frac{\partial^2}{\partial \theta \partial \theta'} \log p(\mathbf{X} | \theta) \quad (\text{A.24})$$

The Langevin diffusion is the following stochastic differential equation:

$$d\theta(t) = \frac{1}{2} \tilde{\nabla}_\theta L\{\theta(t)\} dt + d\mathbf{B}(t) \quad (\text{A.25})$$

where

$$\tilde{\nabla}_\theta L\{\theta(t)\} = -\mathbf{G}^{-1}\{\theta(t)\} \cdot \nabla_\theta L\{\theta(t)\} \quad (\text{A.26})$$

is the gradient of the Riemann manifold associated with the log posterior. The various elements of Brownian motion are defined as follows:

$$\begin{aligned} \mathbf{G}^{-1}\{\theta(t)\} d\mathbf{B}_i(t) = & \left| \mathbf{G}\{\theta(t)\} \right|^{-1/2} \sum_{j=1}^{K_\beta} \frac{\partial}{\partial \theta} \left[\mathbf{G}^{-1}\{\theta(t)\}_{ij} \left| \mathbf{G}\{\theta(t)\} \right|^{1/2} \right] dt \\ & + \left[\sqrt{\mathbf{G}\{\theta(t)\}} d\mathbf{B}(t) \right]_i \end{aligned} \quad (\text{A.27})$$

The discretization of the stochastic differential equation provides a proposal as follows:

$$\begin{aligned} \tilde{\theta}_i = & \theta_i^o + \frac{\varepsilon^2}{2} \left\{ \mathbf{G}^{-1}(\theta^o) \nabla_\theta L(\theta^o) \right\}_i - \varepsilon^2 \sum_{j=1}^{K_\theta} \left\{ \mathbf{G}^{-1}(\theta^o) \frac{\partial \mathbf{G}(\theta^o)}{\partial \theta_j} \mathbf{G}^{-1}(\theta^o) \right\}_{ij} \\ & + \frac{\varepsilon^2}{2} \sum_{j=1}^{K_\theta} \left\{ \mathbf{G}^{-1}(\theta^o) \right\}_{ij} \text{tr} \left\{ \mathbf{G}^{-1}(\theta^o) \frac{\partial \mathbf{G}(\mathbf{a}^o)}{\partial \theta_j} \right\} + \left\{ \varepsilon \sqrt{\mathbf{G}^{-1}(\theta^o)} \xi^o \right\}_i \\ = & \mu(\theta^o, \varepsilon)_i + \left\{ \varepsilon \sqrt{\mathbf{G}^{-1}(\theta^o)} \xi^o \right\}_i \end{aligned} \quad (\text{A.28})$$

where β^o is the current draw. The proposal density is

$$q(\tilde{\theta} | \theta^o) = N_{k_\theta}(\tilde{\theta}, \varepsilon^2 \mathbf{G}^{-1}(\theta^o)) \quad (\text{A.29})$$

and convergence to the posterior distribution is ensured by using the Metropolis-Hastings probability:

$$\min \left\{ 1, \frac{p(\tilde{\theta} | \cdot, Y) q(\theta^o | \tilde{\theta})}{p(\theta^o | \cdot, Y) q(\tilde{\theta} | \mathbf{a}^o)} \right\}. \quad (\text{A.30})$$

Finally, we select ε during the burn-in phase so that, approximately, 25% of all candidates are, eventually, accepted.