Combining Time Series and Cross Sectional Data
for the Analysis of Dynamic Marketing Systems

Csilla Horváth and Jaap E. Wieringa

University of Groningen
Groningen, The Netherlands

SOM-theme F   Interactions between consumers and firms

Abstract

Vector AutoRegressive (VAR) models have become popular in analyzing the behavior of competitive marketing systems. However, an important drawback of VAR models is that the number of parameters to be estimated can become very large. This may cause estimation problems, due to a lack of degrees of freedom. In this paper, we consider a solution to these problems. Instead of using a single time series, we develop pooled models that combine time series data for multiple units (e.g. stores). These approaches increase the number of available observations to a great extent and thereby the efficiency of the parameter estimates. We present a small simulation study that demonstrates this gain in efficiency. An important issue in estimating pooled dynamic models is the heterogeneity among cross sections, since the mean parameter estimates that are obtained by pooling heterogenous cross sections may be biased. In order to avoid these biases, the model should accommodate a sufficient degree of heterogeneity. At the same time, a model that needlessly allows for heterogeneity requires the estimation of extra parameters and hence, reduces efficiency of the parameter estimates. So, a thorough investigation of heterogeneity should precede the choice of the final model. We discuss pooling approaches that accommodate for parameter heterogeneity in different ways and we introduce several tests for investigating cross-sectional heterogeneity that may facilitate this choice. We provide an empirical application using data of the Chicago market of the three largest national brands in the U.S. in the 6.5 oz. tuna fish product category. We determine the appropriate level of pooling and calibrate the pooled VAR model using these data.

Key words: VAR models, Time Series and Cross Sectional data, Pooling

1

1 Introduction

The increased availability of accurate and detailed marketing data has been an important facet in the development and implementation of marketing models. This encourages researchers to propose more complex models with a large(r) number of variables because these models potentially offer better explanations of market phenomena as well as better solutions to economic problems (Leeflang et al. 2000, Leeflang and Wittink 2000).

Vector Autoregressive (VAR) models have been adopted by economists all over the world since the seminal work of Sims (1980) because they are capable of capturing complex dynamic relationships. Recently, with an increasing interest in identifying competitive structures and measuring relationships between marketing variables, VAR models became recognized as effective modelling techniques in marketing. VAR models (i) incorporate all structural relationships in a competitive marketing environment[1] (Dekimpe and Hanssens 1995, Dekimpe et al. 1999, Srinivasan et al. 2001, and Takada and Bass 1998), (ii) treat several variables endogenously (Enders 1995 and Franses 1998), (iii) do not to require firm prior knowledge on the nature of the different relationships (Horváth et al. 2002 and Kornelis 2002), (iv) capture both short- and long-run inter-relationships (Dekimpe and Hanssens 1995, Dekimpe et al. 1999, Horváth et al. 2003, and Kornelis 2002), and (v) outperform univariate time series models in parameter efficiency, goodness-of-fit measures as well as in forecasting performance (Moriarty 1980, Takada and Bass 1998).

However, a major drawback of VAR models is that the number of parameters to be estimated can be very large, depending on the dimension of the system and the order of the model. This may cause problems, due to lack of degrees of freedom available for estimation. The common practise in the marketing literature up until now is to specify VAR models on market-level data and to apply some restrictions on the model, so that the number of parameters to be estimated is reduced.

In this paper we provide a different approach. We specify models at a more disaggregate level and combine the information that is contained in the time series data for multiple units (e.g. supermarket chains, stores, households, or countries) in order to obtain more efficient parameter estimates. We apply four different levels of pooling: we

---

[1]Parsons and Schultz (1976) identify four key elements that need to be incorporated in a parsimonious model for competitive markets: simultaneous relationships, interactions, carryover, and competitive effects. Hanssens (1980) mentions in this respect: sales response effects, competitive reactions, and feedback effects.

consider the Constant Coefficient Model (CCM), the Fixed Effects Model (FEM), the Random Coefficients Model (RCM), and the unit-by-unit approach[2].

The idea of combining the time series and the cross sectional dimensions of the data to arrive at better estimates of a VAR model is not new. In the first Panel VAR (PVAR) model by Holtz-Eakin, Newey and Rosen (1988) the authors formulate a coherent set of procedures for estimating and testing VAR models for panel data. The model is applied to study the dynamic relationships between wages and hours worked in a sample of American males. Another application of this method is a study by Holtz-Eakin et al. (1989) that considers the dynamic relationships between local government revenues and expenditures. Lehr (1999) applies the Holtz-Eakin approach to show that financial intermediation can influence fertility and labor decisions by raising market wages. Binder et al. (2002) suggest a latent variable framework for the analysis of fixed effects PVARs. Several applications build a VAR model from single-equation dynamic panel models. Rousseau and Wachtel (2000) for example, apply the approach of Arellano and Bond (1991) to study dynamic interactions between stock markets and economic performance in individual countries. Canova and Ciccarelli (2000) provide Bayesian methods for forecasting variables and predicting turning points in panel Bayesian VARs.

The references listed above apply models that are developed for *panel data*, i.e. for data with finite T (time span) and large N (number of cross sections). However, many available data sets in marketing have different characteristics, and should be considered in a different setting. In scanner data, for example, the number of cross-sectional units, i.e. stores or products in a category, is usually finite and rather limited while the time span is often quite large (i.e. several weeks). In addition, scanner data is usually only available for specific cross sections (i.e. for a particular store-chain), that cannot be considered as a representative sample of the total population. Hence, one should be reluctant to generalize results of a VAR model that is calibrated on some subset of cross-sectional units to the entire population of all cross-sectional units (e.g. all stores in the Netherlands) but rather interpret the parameter estimates conditional on the observed units. For these reasons, estimation methods that are consistent for large T may be more suitable than PVAR models for marketing applications. Papers that discuss this different setting for estimating pooled univariate time series models are Beck and Katz (1995), Beck (2001), Beck, Epstein, Jackman and O'Halloran (2001), and Beck and Katz (2001). We extend their approach to the case of multiple time series models (i.e. VAR models). In line with the terminology of the above mentioned papers we refer to these estimation procedures as Time Series Cross Sectional (TSCS) methods.

---

[2]Despite the fact that in the unit-by-unit approach, a separate model is specified for each cross-sectional unit, we call it a pooling approach: the unit-by-unit models may be considered as one extreme of the pooling-spectrum.

An additional advantage of estimating VAR models using TSCS procedures is that they do not face the incidental parameter problem that is associated with PVAR models (Neymann and Scott 1948). Moreover, the estimation procedures in a TSCS setting are simpler than the estimation procedures in a PVAR setting.

It may be questionable whether the parameters of economic relationships are stable across cross-sectional units, i.e. there may be heterogeneity in the slope parameters. Hsiao (1986) and Hsiao and Sun (2000) point out that neglecting parameter heterogeneity can be quite serious: "it could lead to inconsistent or meaningless estimates of interesting parameters" (Hsiao 1986, p. 5). Haque, Pesaran and Sharma (2000), Pesaran and Smith (1995), and Robertson and Symons (1992) find that the neglect of slope heterogeneity is yet more serious in dynamic panel data models. The problem arises because in these models the regressors contain lagged endogenous variables. These variables are serially correlated. Therefore, incorrectly ignoring coefficient heterogeneity induces serial correlation in the disturbance term. This generates biased estimates even if $T \rightarrow \infty$. We illustrate these conclusions with a small Monte Carlo simulation study. Our results confirm the findings of the aforementioned authors. Therefore, we argue that it is essential to use a model that accommodates enough heterogeneity so that unbiased estimates of the pooled parameters can be obtained. On the other hand, a model that allows for an unnecessary high degree of heterogeneity requires the estimation of extra parameters and hence, reduces efficiency compared to a model with homogenous parameters. We list and provide a brief discussion of several pooling approaches that accommodate different degrees of heterogeneity and we discuss several pooling tests that facilitate the determination of the appropriate level of cross-sectional heterogeneity.

In summary, the contributions of this paper are:

- we propose to overcome the degrees-of-freedom problem by pooling;

- we provide several pooling approaches that allow for a trade-off between the level of pooling and the degree of heterogeneity;

- we compare the efficiency gains of the different pooling approaches by a small simulation study;

- we propose to estimate the pooled VAR models in a TSCS setting and extend existing univariate TSCS estimation procedures to a multivariate setting;

- we provide guidance on selecting the appropriate level of cross-sectional heterogeneity using pooling tests;

- we illustrate the proposed approaches by an empirical application.

This paper is organized as follows. In Section 2 we specify VAR models for the analysis of pooled data. In Section 3 we discuss four different models that combine the cross-sectional and the time series dimensions of the data. These models accommodate different levels of cross-sectional heterogeneity. The estimation of these models is presented in Section 4. Section 5 compares the small-sample behavior of four main approaches to obtain mean parameter estimates over cross sections in a small Monte Carlo simulation. Guidance on the selection of the appropriate level of heterogeneity is provided Section 6. The proposed approaches are illustrated by an empirical application in Section 7. Finally, we present our conclusions in Section 8.

2 Introduction to VAR models for disaggregate data

The VAR applications in marketing that were referenced in the previous section use a market-level model specification. In this paper, we assume that the available data allow for a model specification on a more disaggregate (store) level. For example, the model that we discuss in Section 7 is specified at the store-level. Furthermore, we assume that T (the time-span) is fairly large.

A general VAR model of order P for cross section i $(i = 1, \cdots, N)$ has the following structure:

$$A_{i,0}Y_{i,t} = \alpha_i + \sum_{t^*=1}^{P} A_{i,t^*} Y_{i,t-t^*} + u_{i,t}, \tag{1}$$

where $Y_{i,t}$ is a k-dimensional vector of endogenous variables of cross section i at time t, A_{i,t^*} is a $k \times k$ matrix that contains the immediate reaction parameters for $t^* = 0$, and the delayed reaction parameters for $t^* \in \{1, 2, \cdots, P\}$ for cross-sectional unit i. The vector α_i is a cross-section-specific intercept, and $u_{i,t}$ is a disturbance term, $u_{i,t} \overset{\text{i.i.d.}}{\sim} N(0, \Sigma_i)$, where Σ_i is usually assumed to be diagonal. This is the structural representation of the VAR model for cross section i.

Multiplication of Equation (1) with any nonsingular $k \times k$ matrix results in an equivalent representation of the process. The reduced form of the model is obtained by pre-multiplying Equation (1) with $A_{i,0}^{-1}$, which gives:

$$Y_{i,t} = \beta_i + \sum_{t^*=1}^{P} C_{i,t^*} Y_{i,t-t^*} + \varepsilon_{i,t}, \tag{2}$$

where $\beta_i = A_{i,0}^{-1}\alpha_i$, $C_{i,t^*} = A_{i,0}^{-1}A_{i,t^*}$ for $t^* = 1, \cdots, P$, and $\varepsilon_{i,t} = A_{i,0}^{-1}u_{i,t}$, with $\varepsilon_{i,t} \overset{\text{i.i.d.}}{\sim} N(0, \Omega_i)$, where $\Omega_i = A_{i,0}^{-1}\Sigma_i A_{i,0}^{-1\prime}$.

5

The model in Equation (2) contains k^2P+k parameters. In the empirical application that we discuss in Section 7 we have $k = 9$ and $P = 2$ so that the model contains 171 VAR-parameters. This large number of parameters puts a rather high demand on the number of available observations for estimation. In addition, VAR estimation procedures rely on asymptotics in T for consistency of the parameter estimates, so that even more observations are needed relative to estimation of the same number of parameters in a standard econometric model. Therefore, a typical problem in applications of VAR models is the shortage of degrees of freedom, and consequently, inefficient estimation results.

Usually, this problem is resolved in marketing applications by simplifying the model in order to allow for more efficient estimation. This has been accomplished by (a) estimating separate models for each brand (Dekimpe et al. 1999, Srinivasan et al. 2000, and Srinivasan and Bass 2001), (b) imposing restrictions on the model based on theory (Horváth et al. 2003), (c) treating some variables as exogenous (Horváth et al. 2003, Nijs et al. 2001, Srinivasan et al. 2000, and Srinivasan and Bass 2001). These solutions impose a priori restrictions on the model. This is not in line with the approach of Sims (1980) who suggests to build a general VAR model where all variables are treated exogenously and use this model to test for exogeneity or for causality. In this paper, we propose to gain degrees of freedom, not by eliminating parameters, but rather by gaining observations through pooling cross-sectional units. In the next section, we discuss four modelling approaches with different degrees of pooling.

3 Four pooling approaches

In this section, we discuss the following four modelling approaches

1. unit-by-unit modelling;

2. constant coefficient modelling;

3. fixed effects modelling;

4. random coefficients modelling.

Ad 1.) The first modelling approach is the collection of models that were discussed in the previous paragraph. Equation (2) specifies separate VAR models for each cross-sectional unit. These models are useful to consider in the ideal situation where the time span of the data is long enough to ensure efficient and reliable estimation of all cross-sectional models, so that there is no need to combine the information across cross sections for better estimation.

In our empirical application the cross sections are stores, so that cross-sectional heterogeneity may result from differences in store profiles (e.g. size of store, shelf space allocated to brands), in consumer profiles (e.g. income, family life cycle), and in competitive profiles (e.g. distance to other stores, number, size, and type of competitive stores).

We refer to the models in Equation (2) as the *unit-by-unit models*. This modelling approach assumes that the cross sections are not related at all and fully accommodates cross-sectional heterogeneity. It can be considered as one extreme of the pooling-spectrum.

Ad 2.) The second modelling approach is at the other extreme, viz. it assumes complete homogeneity across cross sections. We refer to such a model as the *Constant Coefficient Model* (CCM). A CCM version of Equation (2) is:

$$Y_{i,t} = \beta + \sum_{t^*=1}^{P} C_{t^*} Y_{i,t-t^*} + \varepsilon_{i,t} \, , \tag{3}$$

with $\varepsilon_{i,t} \overset{\text{i.i.d.}}{\sim} N(0, \Omega_i)$. The unit-by-unit model and the CCM can be estimated using Feasible Generalized Least Squares (FGLS). Details concerning the estimation of these models can be found in the next section. Relative to the unit-by-unit models, the CCM has more degrees of freedom available for estimation. However, it is at the cost of the often untenable assumption of complete cross-sectional homogeneity.

The third and the fourth modelling approaches can be used to strike a balance between the efficiency of the estimates and the level of heterogeneity that is accommodated by the model.

Ad 3.) For the third modelling approach we assume that the cross-sectional heterogeneity can be captured by cross-section specific intercepts. Specifically, we consider *Fixed Effect Models* (FEMs)[3], where the intercept is cross-section specific, but the other parameters are fully pooled. We specify the FEM analog of the unit-by-unit models in Equation (2) as:

$$Y_{i,t} = \beta_i + \sum_{t^*=1}^{P} C_{t^*} Y_{i,t-t^*} + \varepsilon_{i,t} \, , \tag{4}$$

---

[3]One might also contemplate to consider Random Effect Models (REMs). However, as T grows large, the REM converges to the FEM (see, for example, Beck 2001 or Pesaran and Smith 1995). In addition, the REM can be considered as a special case of RCM. Consequently, here we focus on the FEM and expect similar results for the REM.

where $\varepsilon_{i,t} \overset{\text{i.i.d.}}{\sim} N(0, \Omega)$. In Section 7 we present a VAR model that contains sales response functions, specified in log-log form. For these equations, a store-specific intercept may accommodate sales differences that are due to e.g. store size. If it is reasonable to assume that the customers of the different stores react with the same elasticity to promotional activities, the FEM may capture enough cross-sectional heterogeneity, while preserving a high number of degrees-of-freedom.

Estimation of the parameters of a FEM with FGLS with Dummy Variables (FGLSDV) provides a consistent estimator that are asymptotically efficient for $T \to \infty$ under standard regularity conditions (Bun 2001). This estimation procedure for FEMs is outlined in Section 4.

Ad 4.) The fourth modelling approach allows for cross-sectional heterogeneity in all parameters. However, the restriction is that the parameters vary jointly in a random manner, specified by a certain multivariate distribution. Thus, the parameters of a cross-sectional model are considered as a drawing from some multivariate distribution. These models are referred to as *Random Coefficient Models* (RCMs)[4].

We specify the RCM-analog of Equation (2) as follows[5]:

$$Y_{i,t} = \beta_i + \sum_{t^*=1}^{P} C_{i,t^*} Y_{i,t-t^*} + \varepsilon_{i,t},$$

where $H_i = (\beta_i, C_{i,1}, \cdots, C_{i,P})$, the matrix containing the parameters of the model, satisfies the following restrictions:

$$
\begin{aligned}
\text{vec}(H_i) &\sim N(\text{vec}(H), \Gamma) \\
E(H_i - H | Y_{i,t-1}, Y_{i,t-2}, \cdots, Y_{i,t-P}) &= 0 \\
E(\varepsilon_{i,t} | Y_{i,t-1}, Y_{i,t-2}, \cdots, Y_{i,t-P}) &= 0 \\
E(\varepsilon_{i,t} \varepsilon'_{j,t}) &= \begin{cases} \Omega_i \text{ if } i = j \\ 0 \text{ if } i \neq j. \end{cases}
\end{aligned}
\tag{5}
$$

To get some intuitive understanding of the RCM, it is useful to think of it as a shrinkage estimator. If the individual H_is are completely shrunk back to their mean H, then RCM hardly provides more heterogeneity than the CCM does. On the other hand, if there is almost no shrinkage, then using RCM provides no efficiency gain relative to the GLS estimates of the unit-by-unit models. The degree of shrinkage is a function

---

[4]These models are also often referred to as Random Coefficient Regression Models. The terminology we use is also used by Beck, Epstein, Jackman and O'Halloran (2001) and Beck and Katz (2001).

[5]The methodology described here can be considered as a multiple equation extension of the work of Beck and Katz (2001) and Swamy (1971).

of the heterogeneity in the unit-by-unit estimates and the information that is contained in these estimates. The RCM can be estimated by a modified Swamy estimator. The estimation of RCM, based on the work of Beck and Katz (2001) and Swamy (1971), is also explained in detail in the next section.

4 Estimation

In this section we provide details on the procedures for the estimation of the CCM, the FEM, and the RCM. The unit-by-unit models can be estimated using Feasible Generalized Least Squares (FGLS). FGLS is well-documented in the literature on multiple time series analysis, so that we do not provide details on the estimation of these models here but refer the reader to Lütkepohl (1993) or Hamilton (1994).

4.1 Estimation of the CCM

The relationships in Equation (3) can be rewritten as follows:

$$Y_t = \Phi X_t + \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \cdots + \Pi_P Y_{t-P} + \varepsilon_t, \tag{6}$$

where Y_t is an Nk-dimensional vector that results from stacking the Y_{it} vectors (there are N of them):

$$Y_t = \left(Y'_{1t}, Y'_{2t}, \cdots, Y'_{Nt}\right)'.$$

Furthermore, $\Phi = I_N \otimes B$, where $B = \text{diag}(\beta)^6$, $X_t = \iota_{Nk}$, with ι_{Nk} an Nk-dimensional vector of ones[7], $\Pi_j = I_N \otimes C_j$ is a block diagonal $Nk \times Nk$-matrix, and ε_t is an Nk-dimensional disturbance vector:

$$\varepsilon_t = \left(\varepsilon'_{1t}, \varepsilon'_{2t}, \cdots, \varepsilon'_{Nt}\right)',$$

---

[6]Where we use the notation

$$\text{diag}(\beta) = \begin{pmatrix} \beta_{(1)} & & 0 \\ & \ddots & \\ 0 & & \beta_{(k)} \end{pmatrix},$$

where $\beta_{(j)}$ denotes the jth element of β.

[7]The estimation procedures that are outlined in this section are easily extended to the case of VARX modelling by including exogenous variables in the X_t vector. However, in that case B does not equal $\text{diag}(\beta)$, see also footnote 9.

that is normally distributed with expectation 0 and variance-covariance matrix

$$
\begin{pmatrix}
\Omega_1 & & 0 \\
& \ddots & \\
0 & & \Omega_N
\end{pmatrix}.
$$

Let us rewrite Equation (6) in the following way[8]:

$$
\begin{aligned}
Y_t &= \text{vec}(Y_t) \\
&= \text{vec}\left[(I_N \otimes B)X_t + (I_N \otimes C_1)Y_{t-1} + \cdots + (I_N \otimes C_P)Y_{t-P} + \varepsilon_t\right] \\
&= \text{vec}\left(BX_t^*\right) + \text{vec}\left(C_1 Y_{t-1}^*\right) + \cdots + \text{vec}\left(C_p Y_{t-P}^*\right) + \varepsilon_t \\
&= \text{vec}\left[(L)\begin{pmatrix} X_t^* \\ Y_{t-1}^* \\ \vdots \\ Y_{t-P}^* \end{pmatrix}\right] + \varepsilon_t \\
&= \left((X_t^{*\prime}, Y_{t-1}^{*\prime}, \cdots, Y_{t-P}^{*\prime}) \otimes I_k\right)\text{vec}(L) + \varepsilon_t{}^9
\end{aligned} \tag{7}
$$

where

$$
\begin{aligned}
Y_t^* &= (Y_{1t}, Y_{2t}, \cdots, Y_{Nt}), \\
X_t^* &= (X_{1t}, X_{2t}, \cdots, X_{Nt}), \text{ and} \\
L &= (B, C_1, \cdots, C_P).
\end{aligned}
$$

From Equation (7), it follows that we can estimate the CCM-VAR model by a simple reparameterization. Define

$$
Y = \left(Y_{p+1}^{*\prime}, Y_{p+2}^{*\prime}, \cdots, Y_T^{*\prime}\right)',
$$

---

[8]We use the following property of the vec-operator several times: $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$, which holds for matrices A, B and C that are of such dimensions that the matrix product ABC is defined (see Magnus and Neudecker (1988) page 31, Equation (7).

[9]Note that Equation (7) can be simplified by noticing that

$$
\begin{aligned}
\left(X_t^{*\prime} \otimes I_k\right)\text{vec}(\text{diag}(\beta)) &= \text{vec}\left(\text{diag}(\beta)\iota_k \iota_N'\right) \\
&= \text{vec}\left(\beta \iota_N'\right) \\
&= (\iota_N \otimes I_k)\beta.
\end{aligned}
$$

Hence, in Equation (7), $X_t^{*\prime}$ can be replaced by ι_N, and the first k^2 elements of $\text{vec}(L)$ by β. We opt for this notation because the formulas above are also valid for estimating a VARX model. In that case, observations of the eXogenous variables are added to the $X_{i,t}$ vector, so that its length increases, and the number of columns of B increases accordingly. Note that in the VARX case, the simplification in this footnote also holds, but then only for that part of the $X_t^{*\prime}$-matrix that deals with the intercepts.

and

$$W = \begin{pmatrix} (X^{*'}_{p+1}, Y^{*'}_{p}, \cdots, Y^{*'}_{1}) & \otimes I_k \\ (X^{*'}_{p+2}, Y^{*'}_{p+1}, \cdots, Y^{*'}_{2}) & \otimes I_k \\ \vdots \\ (X^{*'}_{T}, Y^{*'}_{T-1}, \cdots, Y^{*'}_{T-p}) \otimes I_k \end{pmatrix} = \begin{pmatrix} W_p \\ W_{p+1} \\ \vdots \\ W_{T-1} \end{pmatrix}, \tag{8}$$

and finally,

$$\varepsilon = \left(\varepsilon'_{p+1}, \varepsilon'_{p+2}, \cdots, \varepsilon'_T \right)'.$$

In this notation, the model is

$$Y = W \text{vec}(L) + \varepsilon. \tag{9}$$

The OLS estimate for the parameter vector $\text{vec}(L)$ is obtained by

$$\widehat{\text{vec}(L)} = (W'W)^{-1} W'Y. \tag{10}$$

The disturbances of the reduced VAR system are, in general, contemporaneously correlated. Hence, a system estimator is applicable. Zellner's (1962) Estimated Generalized Least Squares (EGLS) estimator can be used to gain efficiency from the cross-equation correlations of the OLS disturbances[10]:

$$\widetilde{\text{vec}(L)} = (W' \widehat{\Psi}^{-1} W)^{-1} W' \widehat{\Psi}^{-1} Y, \tag{11}$$

---

[10]In case of panel homoscedasticity the GLS estimator of an unrestricted VAR system reduces to the OLS estimator. This can be proven as follows. Rewrite Equation (8) in the following way:

$$W = D \otimes I_k, \quad \text{where } D = \begin{pmatrix} X^{*'}_{p+1} & Y^{*'}_{p} & \cdots & Y^{*'}_{1} \\ X^{*'}_{p+2} & Y^{*'}_{p+1} & \cdots & Y^{*'}_{2} \\ \vdots & \vdots & & \vdots \\ X^{*'}_{T} & Y^{*'}_{T-1} & \cdots & Y^{*'}_{T-p} \end{pmatrix}.$$

The OLS estimator of $\text{vec}(L)$ is

$$\widehat{\text{vec}(L)}_{\text{OLS}} = \left((D \otimes I_k)'(D \otimes I_k) \right)^{-1} (D \otimes I_k)' \text{vec}(Y).$$

The GLS estimator of $\text{vec}(L)$ is

$$\begin{aligned}
\widehat{\text{vec}(L)}_{\text{GLS}} &= \left((D \otimes I_k)' \left(I_{N(T-p)} \otimes \Omega \right)^{-1} (D \otimes I_k) \right)^{-1} (D \otimes I_k)' \left(I_{N(T-p)} \otimes \Omega \right)^{-1} \text{vec}(Y) \\
&= \left(D'D \otimes \Omega^{-1} \right)^{-1} \left(D' \otimes \Omega^{-1} \right) \text{vec}(Y) \\
&= \left(\left(I_{k(p+1)} D'D \right) \otimes \left(\Omega^{-1} I_k \right) \right)^{-1} \left(I_{k(p+1)} D' \right) \otimes \left(\Omega^{-1} I_k \right) \text{vec}(Y) \\
&= \left(D'D \otimes I_k \right)^{-1} \left(I_{k(p+1)} \otimes \Omega^{-1} \right)^{-1} \left(I_{k(p+1)} \otimes \Omega^{-1} \right) \left(D' \otimes I_k \right) \text{vec}(Y) \\
&= \left((D \otimes I_k)'(D \otimes I_k) \right)^{-1} (D \otimes I_k)' \text{vec}(Y).
\end{aligned}$$

□

11

where

$$\Psi = I_{(T-p)} \otimes \begin{pmatrix} \Omega_1 & & 0 \\ & \ddots & \\ 0 & & \Omega_N \end{pmatrix}.$$

4.2 Estimation of the FEM

For estimating the FEM, we use FGLS with Dummy Variables (FGLSDV). The estimation procedure proceeds analogously to that of the CCM[11]. The differences are that in Equation (6), we now define Φ as:

$$\Phi = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_N \end{pmatrix}$$

where $B_i = \text{diag}(\beta_i)$. After applying the simplification of footnote 9, we have in Equation (7) that $X_t^* = I_N$ and the first kN elements of $\text{vec}(L)$ are $(\beta_1', \beta_2', \cdots, \beta_N')'$.

4.3 Estimation of the RCM

Let us rewrite the VAR-model in a compact form:

$$Y_i = H_i Z_i + \varepsilon_i, \tag{12}$$

where $Y_i = (Y_{i,1}, ..., Y_{i,T})$ and H_i was defined in Section 3. Furthermore, $Z_i = (Z_{i,P+1}, ..., Z_{i,T})$, with $Z_{i,t} = (X_{i,t}', Y_{i,t-1}', Y_{i,t-2}', \cdots, Y_{i,t-P}')'$, where $X_{i,t} = 1$[12]. Finally, $\varepsilon_i = (\varepsilon_{i,1}, ..., \varepsilon_{i,T})$ where $\varepsilon_{i,t} \overset{\text{i.i.d.}}{\sim} N(0, \Omega_i)$.

Next, we define:

$$G_i = H_i - H , \tag{13}$$

so that $\text{vec}(G_i) \sim N(0, \Gamma)$. Using this in Equation (12) results in:

$$\begin{aligned} Y_i &= H Z_i + (\varepsilon_i + G_i Z_i) \\ Y_i &= H Z_i + \omega_i, \end{aligned} \tag{14}$$

---

[11] Analogous to the proof for the CCM it is possible to show that, in case of panel homoscedasticity, GLS with dummy variables for an unrestricted VAR system reduces to OLS with dummy variables.

[12] This specification is extended to a VARX model in a straightforward manner by redefining $X_{i,t}$ as a vector whose first element equals one, and the remaining elements are the values of the eXogenous variables at time t.

where ω_i is the new composite error term. The first part of this error term (ε_i) is the standard stochastic part of the regression model. The second part ($G_i Z_i$) is the error associated with the deviation of the H_i of a particular cross-sectional unit i from the overall mean H. Vectorizing Equation (14) gives:

$$\text{vec}(Y_i) = (Z_i' \otimes I_k)\text{vec}(H) + \text{vec}(\omega_i).$$

The covariance matrix of $\text{vec}(\omega_i)$ is:

$$
\begin{aligned}
& E\left[\text{vec}(\omega_i)\text{vec}(\omega_i)'\right] \\
={} & E\left[\left(\text{vec}(\varepsilon_i) + (Z_i' \otimes I_k)\text{vec}(G_i)\right)\left(\text{vec}(\varepsilon_i) + (Z_i' \otimes I_k)\text{vec}(G_i)\right)'\right] \\
={} & I_T \otimes \Omega_i + (Z_i' \otimes I_k)\Gamma(Z_i \otimes I_k) = \Upsilon_i.
\end{aligned}
$$

Furthermore, we assume that the errors are not correlated across cross sections, so that Λ, the covariance matrix of $\omega = (\omega_1, \omega_2, \cdots, \omega_N)$, has the following structure:

$$
E\left[\text{vec}(\omega)\text{vec}(\omega)'\right] = \Lambda = \begin{pmatrix} \Upsilon_1 & 0 & \cdots & 0 \\ 0 & \Upsilon_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Upsilon_N \end{pmatrix}.
$$

We employ GLS to obtain consistent estimates of H and H_i. In the Appendix we show that h, the GLS estimator of $\text{vec}(H)$, is a weighted average of h_i^{OLS}, the unit-by-unit OLS estimates of the $\text{vec}(H_i)$'s:

$$h = \sum_{i=1}^{N} W_i\, h_i^{\text{OLS}}, \tag{15}$$

where the weights can be written as:

$$W_i = \left[\sum_{i=1}^{N}(V_i + \Gamma)^{-1}\right]^{-1}(V_i + \Gamma)^{-1},$$

with

$$V_i = (Z_i Z_i')^{-1} \otimes \Omega_i.$$

The weights are inversely related to the variance of the unit-by-unit OLS estimates since:

$$\text{Var}\left(h_i^{\text{OLS}}\right) = (Z_i Z_i' \otimes I_k)^{-1} \left[(Z_i \otimes I_k)(I_T \otimes \Omega_i)(Z_i' \otimes I_k)\right](Z_i Z_i' \otimes I_k)^{-1}$$
$$+ (Z_i Z_i' \otimes I_k)^{-1}(Z_i Z_i' \otimes I_k)\Gamma(Z_i Z_i' \otimes I_k)(Z_i Z_i' \otimes I_k)^{-1}$$
$$= V_i + \Gamma.$$

For the variance of h we obtain:

$$\text{Var}(h) = \sum_{i=1}^{N} W_i \text{Var}(h_i^{\text{OLS}}) W_i'$$
$$= \sum_{i=1}^{N} W_i \left[V_i + \Gamma\right] W_i'.$$

Let us denote the Best Linear Unbiased Predictor (BLUP) of $\text{vec}(H_i)$ that comes from the GLS framework by h_i. For h_i we have the following[13]:

$$h_i = (\Gamma^{-1} + V_i^{-1})^{-1}(\Gamma^{-1}h + V_i^{-1}h_i^{\text{OLS}})$$
$$= T_i h + (I_Q - T_i)h_i^{\text{OLS}},$$

where Q is the number of regressors in the VAR model[14] and $T_i = \left[\Gamma^{-1} + V_i^{-1}\right]^{-1}\Gamma^{-1}$. So h_i is the weighted average of the estimates of the most and least pooled model. Its variance equals:

$$\text{Var}(h_i) = T_i \text{Var}(h)T_i' + (I_Q - T_i)\text{Var}(h_i^{\text{OLS}})(I_Q - T_i)' +$$
$$(I_Q - T_i)\text{Cov}(h, h_i^{\text{OLS}})T_i' + T_i \text{Cov}(h, h_i^{\text{OLS}})(I_Q - T_i)',$$

where

$$\text{Cov}(h, h_i^{\text{OLS}}) = W_i \text{Var}(h_i^{\text{OLS}})$$
$$= \left[\sum_{i=1}^{N}(V_i + \Gamma)^{-1}\right]^{-1}.$$

---

[13] See Judge et al. (1985), pp. 541.

[14] An important assumption for the calculations of H_is of Swamy (1971) is that $N > Q$, i.e. that the number of cross sections is higher than the number of parameters of the model. This is important because otherwise Γ will not be of full rank. It will be singular and hence, one cannot take its inverse and H_is cannot be calculated. In the case of a VARX model Q may be quite large. Therefore, the assumption may not hold with the TSCS data where one assumes finite (few) N and large T. In this case one could save significant number of parameters, and hence regain non-singularity of Γ, by imposing some structure on it, for example, by restricting it to be diagonal.

The remaining problem is to estimate V_i and Γ. We estimate V_i by:

$$\widehat{V}_i = \left[(Z_i Z_i')^{-1} \otimes \widehat{\Omega}_i \right],$$

where

$$\widehat{\Omega}_i = \frac{e_i e_i'}{T - Q/k},$$

where e_i are the standard OLS residuals.

For estimating Γ, one would like to observe the individual $\text{vec}(H_i)$s. Instead, we only have noisy estimates in the form of h_i^{OLS} available. The Swamy estimator (Swamy 1971) corrects for this extra sampling variability. The problem with this estimator is that in finite samples it may not provide a positive definite $\widehat{\Gamma}$. In order to solve this problem we apply the approach of the so-called 'Beck-Katz kludge' (Beck and Katz (2001)):

$$\widehat{\Gamma} = \max \left[0, \frac{1}{N-1} \left(\sum_{i=1}^{N} h_i^{\text{OLS}} h_i^{\text{OLS}'} - N \overline{h_i^{\text{OLS}}} \, \overline{h_i^{\text{OLS}}}' \right) - \frac{1}{N} \left(\sum_{i=1}^{N} \widehat{V}_i \right) \right],$$

where $\overline{h_i^{\text{OLS}}}$ is the mean of the h_i^{OLS}s. The resulting estimate for Γ is equal to the Swamy estimator (right expression in the square brackets) if this is a positive definite matrix, whereas $\widehat{\Gamma} = 0$ if the Swamy estimator is not positive definite. In that case, the RCM estimates of $\text{vec}(H_i)$ are equal to the CCM estimate.

The modelling approaches that are discussed in this and in the previous section allow for different levels of cross-sectional heterogeneity. To obtain some insights in the biases that arise when applying the different approaches to estimate mean parameters we run a small Monte Carlo simulation study that we present in the next section. In this study we compare the four most widely used approaches; (1) estimating a VAR model from aggregate data, (2) averaging disaggregate estimates (unit-by-unit model), (3) FEM, and (4) RCM. This means that in the simulation study we also consider the approach that is usually employed in marketing applications, viz. estimating a VAR model from aggregate data. In addition, we do not discuss the CCM here since it can be considered as a special case of the FEM and the main arguments that hold for the FEM are also valid for the CCM.

5   A comparison of four ways to obtain mean parameter estimates

In this section, we present a small Monte Carlo simulation study to address the small-sample properties of the the four most widely used approaches that provide

mean parameter estimates. We generate 1000 data-sets from VAR(1) models with two endogenous variables and a constant intercept term for 10 cross-sectional units over 100 time periods. We draw the error terms from a multivariate normal distribution with zero expected value. To induce heterogeneity between the cross sections we split them into two homogenous groups of 5. We induce different data-generating processes (i.e. different VAR parameters) for the groups. The original parameter values of the two models are presented in Table 1 together with other details concerning the setup of the simulation study[15]. We apply the four approaches to estimate the mean parameters of the VAR model that were discussed in the previous section: (1) we aggregate the data over the 10 cross sections and estimate a VAR model from the aggregated data (aggregate approach), (2) we build separate VAR models for each cross section and average the estimated parameters afterwards (averaged unit-by-unit approach), (3) we estimate mean parameters from a FEM, and (4) we estimate mean parameters using an RCM.

Table 1: Summary statistics of the simulation study

Number of draws			1000			
Number of data points (T)			100			
Number of cross sections (N)			10			
	Slope Parameters				*Constants*	
Original 1 (5 units)	-0.3	0.1	0.1	-0.4	2	3
Original 2 (5 units)	0.7	-0.5	0.3	0.5	2	3
Original mean values	**0.2**	**-0.2**	**0.2**	**0.05**	**2**	**3**
Var-cov matrix of the residuals	1	0.3	0.3	0.7		

The simulated distributions of the VAR parameter estimates for each of the different estimation techniques are visualized in the box-plots of Figure 1. The dashed line indicates the original (mean) parameters. Each subplot in the figure shows the box-plots of the simulated parameter estimates for the four approaches of one of the four

---

[15]The parameter matrices of the VAR system are vectorized and transposed in Table 1. Hence, the VAR model for the second group is:

$$Y_t = \begin{pmatrix} 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 0.7 & 0.3 \\ -0.5 & 0.5 \end{pmatrix} Y_{t-1} + \varepsilon_t \text{ , with } \mathrm{Cov}(\varepsilon_t) = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}.$$

parameters. The upper-left subplot shows, for example, the box-plots of the simulation results for the first parameter in Table 1. The upper-right subplot corresponds to the second, the lower left to the third, and the lower right to the fourth parameter. The

Figure 1: Box-plots of the simulated parameter estimates
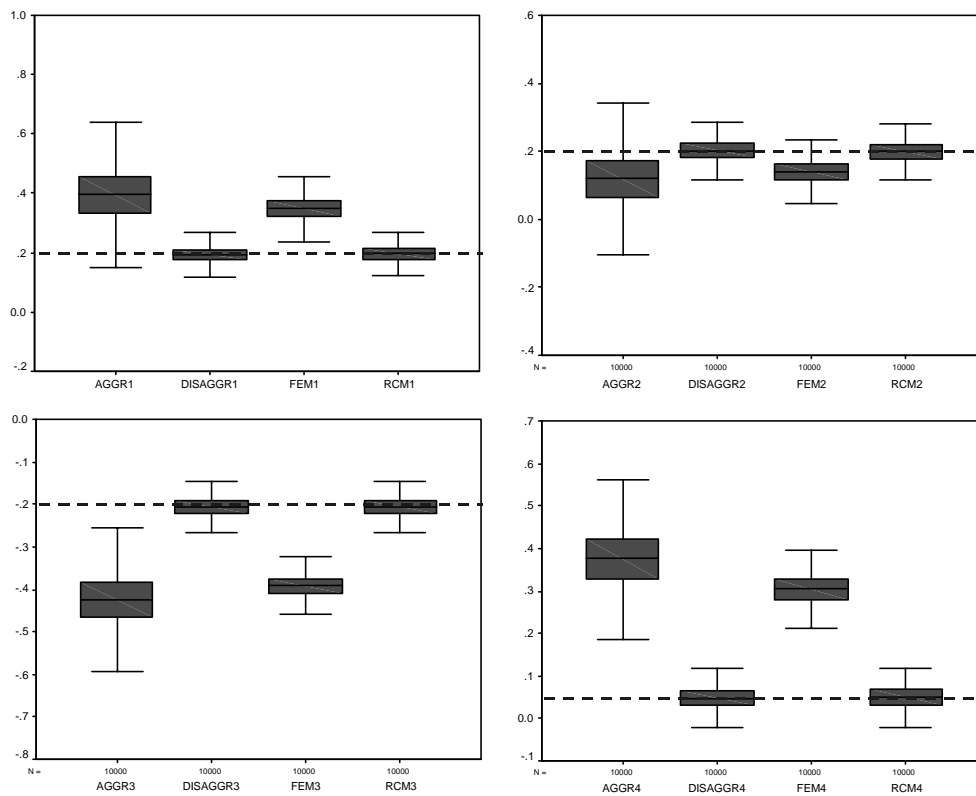
The dashed lines indicate the original (mean) parameters.

box-plots clearly indicate that (for the setup of the simulation study) aggregation and the FEM approach (i.e. the approaches that do not allow for heterogeneity in the slope coefficients) lead to biased estimates of the mean parameters. The RCM performs quite well (its mean and median are the closest to the original mean and it has relatively low standard deviation of the parameter estimates) despite the fact that the distribution of the parameters over the cross sections is far from normal. The disaggregate approach also performs well: its results are comparable to the RCM. To gain further insight into the small sample behavior of the different approaches we compute the Average

17

Squared Prediction Error (ASPE), also known as the Mean Squared Error (MSE), for each approach. The ASPE captures both the bias and the variance of the parameter estimates. We decompose it according to Leeflang et al. (2000, p. 507) into Average Prediction Error (APE) that captures bias and into a variance component that captures unreliability (variance), see Table 2. We see that the RCM and the disaggregate

Table 2: Monte Carlo experiment results - experiment 1

	Slope Parameters				Constants	
ASPE (MSE)						
Aggregate	0.0462	0.0132	0.0542	0.1091	0.2805	0.5379
FEM	0.0233	0.0050	0.0371	0.0654	1.4581	0.5835
Disaggregate	0.0008	0.0010	0.0006	0.0007	0.0178	0.0108
RCM	0.0008	0.0006	0.0010	0.0007	0.0163	0.0102
APE						
Aggregate	0.1943	-0.0811	-0.2244	0.3229	0.3246	-0.6798
FEM	0.1471	-0.0606	-0.1910	0.2534	-1.2006	-0.7564
Disaggregate	-0.0076	0.0017	-0.0061	-0.0019	0.0577	0.0296
RCM	-0.0041	-0.0066	-0.0006	-0.0002	0.0449	0.0220
Variance						
Aggregate	0.0085	0.0066	0.0039	0.0048	0.1751	0.0758
FEM	0.0017	0.0013	0.0006	0.0012	0.0166	0.0113
Disaggregate	0.0008	0.0010	0.0005	0.0007	0.0145	0.0099
RCM	0.0008	0.0005	0.0010	0.0007	0.0143	0.0097

approach outperform the aggregate approach and the FEM both in terms of variability and unbiasedness. RCM performs slightly better than the disaggregate approach especially with respect to variance. This lower variance is due to the efficiency gain from combining cross-sectional units. For the slope parameters, the aggregate approach is the worst according to all measures. For the constants, the disaggregate approach and the RCM approach perform better than the aggregate approach and the FEM approach. The poor performance of the aggregate approach is due to both larger bias and larger variability, whereas the poor performance of FEM is due to larger bias only.

To examine the relative performance of the approaches when the number of observations is smaller, we run new Monte Carlo experiments with a smaller number of time points ($T = 10$). We decompose the ASPE analogously to Table 2. The results are presented in Table 3. We see that all methods perform worse than in the

Table 3: Monte Carlo experiment results - experiment 2

	Slope Parameters				Constants	
ASPE (MSE)						
Aggregate	0.1545	0.0482	0.1081	0.1689	1.2335	0.8694
FEM	0.0388	0.0217	0.1024	0.1014	1.5770	0.6120
Disaggregate	0.0136	0.0182	0.0132	0.0108	0.3259	0.2105
RCM	0.0108	0.0163	0.0115	0.0095	0.2753	0.1801
APE						
Aggregate	0.3337	-0.0635	-0.2893	0.3780	-0.2041	-0.4514
FEM	0.1679	-0.0932	-0.3123	0.3058	-1.1796	-0.6944
Disaggregate	-0.0598	-0.0290	-0.0578	-0.0411	0.3916	0.3218
RCM	-0.0359	-0.0333	-0.0560	-0.0293	0.3303	0.2735
Variance						
Aggregate	0.0432	0.0442	0.0244	0.0261	1.1918	0.6656
FEM	0.0106	0.0130	0.0049	0.0079	0.1857	0.1298
Disaggregate	0.0101	0.0174	0.0099	0.0091	0.1725	0.1070
RCM	0.0095	0.0152	0.0084	0.0086	0.1662	0.1053

previous case due to the small sample bias that gets more serious for short time series and due to the reduced number of degrees of freedom. Again, the approaches that allow for heterogeneity in the slope parameters outperform the other two. The aggregate method has the highest ASPE values and the highest variance. The variances of the parameters using the other three estimators increase as the sample size gets smaller but they all stay relatively small. The FEM estimates have in general the smallest variance. This is probably due to the fact that this model applies the strongest assumption about cross-sectional heterogeneity and hence, has the largest number of degrees of freedom. This becomes essential when the number of available longitudinal observations is small. However, the estimates are significantly more biased than those of the unit-by-unit model or the RCM. The last two approaches perform very similarly with respect to bias but the RCM parameters have smaller variation due to the efficiency gain from combining the cross sections. This results in lower ASPE values, too.

Our findings are in line with those of Pesaran and Smith (1995) who examine the consistency of several approaches to estimate the average coefficients in dynamic models. They find that "when the coefficients differ across groups (cross sections), pooling and aggregating give inconsistent and potentially highly misleading estimates

19

of the coefficients ... in dynamic models, even for large N and T"

Bartels (1996) argues that "we are always in the position of deciding how much we should pool some observations with others, and we always have a choice ranging from complete pooling to assuming that the data have nothing to do with each other". He proposes that one should estimate a model allowing for varying degrees of pooling and then take a scientific decision after examining the locus of all such estimates. In the next section we discuss several test that facilitate such a decision.

6   Testing poolability

Despite the widespread application of panel and TSCS data models in practise, testing of the so-called poolability hypothesis of constant slope coefficients is often neglected in empirical analysis. However, in the previous section we demonstrated that, when estimating a regression relationship, poor inference may be drawn if the true slope coefficients vary across units and we do not accommodate for this heterogeneity, especially when dynamics are present (see also Bun 2001, Pesaran and Smith 1995, and Robertson and Symons 1992). As we explained before, the problem arises because in a dynamic model the regressors are serially correlated. Incorrectly ignoring coefficient heterogeneity induces serial correlation in the disturbance term. This generates inconsistent estimates in models with lagged endogenous variables.

These findings suggest that unbiased estimation of the mean parameters of dynamic models requires allowing for a sufficiently high degree of cross-sectional heterogeneity. At the same time, a model that needlessly allows for heterogeneity requires the estimation of extra parameters and hence, reduces efficiency of the parameter estimates. So, a thorough investigation of heterogeneity should precede the choice of the final TSCS model. To facilitate this investigation, we briefly discuss several pooling tests in this section.

The most popular test for the poolability of TSCS data is the Chow (1960) F-test. This test can be applied to investigate the null-hypothesis of parameter homogeneity: $H_i = H_j$ for all $i \neq j$[16]. The statistic that follows an F-distribution with r and d degrees of freedom is:

$$F = \frac{(\text{RRSS} - \text{URSS})/r}{\text{URSS}/d},$$

where RRSS is the sum of squared residuals from the restricted (pooled) model, URSS the sum of squared residuals from the unrestricted (unit-by-unit) model, r is the number

---

[16]The Chow test can also be applied on a subset of the model parameters. It can, for example be used to evaluate the null hypothesis of equal slope coefficients for the cross sections, but different intercepts.

of linearly independent restrictions, and d is the number of degrees of freedom for the unrestricted model. This statistic has a logical extension for the system of equation context, viz. instead of using sums of squared residuals, we use the trace of the variance-covariance matrix of residuals.

Swamy (1971), p. 123–124, suggested that a test of the random coefficient model can be based on the differences between the unit-by-unit estimates and a weighted average of these estimates:

$$\chi^2 = \sum_{i=1}^{N} \left(h_i^{\text{OLS}} - \widetilde{h} \right)' \widehat{V}_i^{-1} \left(h_i^{\text{OLS}} - \widetilde{h} \right),$$

where

$$\widetilde{h} = \left(\sum_{i=1}^{N} \widehat{V}_i^{-1} \right)^{-1} \sum_{i=1}^{N} \widehat{V}_i^{-1} h_i^{\text{OLS}},$$

where h_i^{OLS} and $\widehat{V}_i$ are defined as in Section 4. This statistic follows a χ^2-distribution with $Q(N-1)$ degrees of freedom where Q is the number of parameters for one cross section in the VAR model, and N is the number of cross sections.

Bun (2001b) examines the small sample performance of these classical asymptotic tests for regression models with both lagged dependent variables and nonspherical disturbances. His conclusion is that these test procedures perform rather poorly when using critical values from either the F- or the χ^2- distribution. He suggests to use the original Chow F-test statistics with bootstrapped critical values to increase accuracy.

Baltagi, Hidalgo and Li (1996) derive a nonparametric poolability test for panel data. They find that the advantage of this approach over conventional parametric tests is its robustness to regression functional form misspecification.

Another way of exploring heterogeneity is cross-validation. Beck (2001) suggests this approach for TSCS data. This procedure involves estimating the model several times, leaving one cross-sectional unit out at a time. The estimation results are compared to find out whether some cross sections are predicted less well than others.

In the empirical application that we discuss in the next section, we apply the Chow-test to test the poolability of the parameters. To this end, we start by estimating the unit-by-unit models. Subsequently, we estimate the three remaining models and test whether pooling is allowed.

7 Empirical Application

We study the Chicago market of the three largest national brands in the U.S. in the 6.5 oz. tuna fish product category. We have 104 weeks of observations for each of 28

stores of one supermarket chain in the metropolitan area, covering unit sales, actual and regular prices, features, and displays. This is the data set that was used in the study of Horváth et al. (2003). We use the data pooled for 26 stores[17].

7.1 Model Specification

The variables of interest are the logarithms of unit sales and the logarithms of price indices (ratio of actual to regular price) of three brands at the store level. We define two types of price promotion variables: (1) own- and other-brand temporary discounts without support and (2) own- and other-brand temporary discounts with feature only, display only, or feature and display support. Van Heerde et al. (2000, 2001) use this approach to allow for interaction effects between discounts and support. Also, the promotion variables are minimally correlated by definition[18]. We treat the sales variables and the price variables of the 3 brands endogenously in the system of equations. We consider the non-price instruments to be exogenous to the model (hence, in fact, we build a VARX model). Furthermore, we do not include lagged non-price instruments but we allow for dynamic effects indirectly through the inclusion of lagged endogenous variables. For each brand, we include a sales response function, and two competitive reaction functions (one for supported price and one for non-supported price). This results in a 9-dimensional VARX model with three sales response functions and six competitive reaction functions.

In Section 3 we discussed several alternative pooling approaches. In order to save space, we do not present the four different versions of the VARX model in this section. We only show the FEM version because, as we will see later in this section, this is the appropriate pooling approach for our data. The specifications of the other pooling approaches follow from Equations (16) and (17) that are presented below in a straightforward manner.

We specify the sales response functions for the FEM-VARX model as follows:

---

[17]In two stores the brands are not promoted. We exclude these stores from the analysis.

[18]Van Heerde et al. (2000, 2001) use four different price promotion variables. We use only two variables to reduce degrees-of-freedom problems.

$$\ln S_{qi,t} = \alpha_{qi} + \sum_{k=1}^{2} \sum_{t^*=0}^{P} \sum_{j=1}^{n} \beta_{PIijk,t^*} \ln PI_{qjk,t-t^*}$$

$$+ \sum_{j=1}^{n} \sum_{t^*=1}^{P} \varphi_{ij,t^*} \ln S_{qj,t-t^*} + \sum_{j=1}^{n} \beta_{Fij} F_{qj,t} \qquad (16)$$

$$+ \sum_{j=1}^{n} \beta_{Dij} D_{qj,t} + \sum_{j=1}^{n} \beta_{FDij} FD_{qj,t} + \varepsilon_{qi,t}$$

$$(q = 1, \cdots, N, \ i = 1, \cdots, n, \ \text{and} \ t = 1, \cdots, T),$$

where

$\ln S_{qi,t}$ is the natural logarithm of sales of brand i in store q in week t;

$\ln PI_{qik,t}$ is the log price index (actual to regular price) of brand i in store q in week t ($k = 1$ denotes prices that are supported and $k = 2$ denotes prices that are not supported);

$F_{qj,t}$ is the feature-only dummy variable for non-price promotion of brand j in store q at time t;

$D_{qj,t}$ is the display-only dummy variable for non-price promotion of brand j in store q at time t;

$FD_{qj,t}$ is the feature-and-display dummy variable for non-price promotion of brand j in store q at time t;

α_{qi} is a store specific intercept for the equation corresponding to brand i and store q;

β_{PIijk,t^*} is the pooled elasticity of brand i's sales with respect to brand j's price index;

φ_{ij,t^*} is the pooled elasticity of brand i's sales with respect to its own sales in week $t - t^*$;

β_{Fij}, β_{Dij}, and β_{FDij} are the pooled current-week effects on brand i's log sales resulting from brand j's use of feature-only (F), display-only (D), and feature and display (FD);

n is the number of brands in the product category;

N is the number of stores;

23

$\varepsilon_{qi,t}$ are the disturbances.

Note that Equation (16) does not contain cross-sectional unit effects: i.e. we do not assume that promotions in other stores $q' \neq q$, $q' = 1, \cdots, N$, have an effect on $S_{qi,t}$.

Using the same notation, we specify the competitive reaction functions for the FEM as follows:

$$
\begin{aligned}
\ln PI_{qil,t} \quad = \quad & \delta_{qil} + \sum_{t^*=1}^{P} \gamma_{il,t^*} \ln PI_{qil,t-t^*} + \sum_{\substack{t^*=1 \\ k \neq l}}^{P} \gamma_{ik,t^*} \ln PI_{qik,t-t^*} \\
& + \sum_{k=1}^{2} \sum_{t^*=1}^{P} \sum_{\substack{j=1 \\ j \neq i}}^{n} \gamma_{iljk,t^*} \ln PI_{qjk,t-t^*} \qquad (17) \\
& + \sum_{t^*=0}^{P} \sum_{j=1}^{n} \eta_{ij,t^*} \ln S_{qj,t-t^*} + \upsilon_{qit} \\
(q \quad = \quad & 1, \cdots, N, \quad i = 1, \cdots, n \quad l = 1, 2 \text{ and } t = 1, ..., T) \,,
\end{aligned}
$$

where all variables are defined as in Equation (16).

The VARX model captures several dynamic phenomena; competitive reactions, own- and cross-feedback, intrafirm effects (relations between different variables for the same brand), inertia (lagged endogenous effects), purchase reinforcement, immediate sales response, and delayed response.

7.2 Testing poolability

We first consider the FEM calibrated for tuna data. Based on the Schwartz Criterion (SC) we set the order of the VARX model equal to two[19]. We test heterogeneity/homogeneity of the coefficients across cross-sectional units twice.

Firstly, we perform a series of Chow tests. To test for overall parameter homogeneity, we compare the sum of squared residuals (SSR) of the unit-by-unit

---

[19]We compute the SC's for FEMs with order 1 to 12 to determine the optimal order of the model for the tuna data. We find that SC decreases with the inclusion of extra lags. This is probably due to the fact that with many cross sections additional lags are easy to accommodate. Specifically, we find that the part that captures the model fit in SC gets substantially higher weight than the part that penalizes for loss in degrees of freedom. In addition, we find that the model of order 12, for example, suffers from a high amount of multicollinearity. To overcome this problem, we compute the SC's on randomly selected sets of 10 stores, and find the order is between 2 and 4. Importantly, the effects are robust across models with orders varying from 2 to 6.

model to that of the CCM. To test for slope homogeneity, we perform Chow test for deciding between the unit-by-unit model and the FEM. Finally, to test for homogeneous intercepts conditional on the acceptance of homogeneous slopes, we compare the SSR of the FEM to that of the CCM. We present the test statistics of these three Chow tests in Table 4[20]. In all cases we accept the null hypothesis of homogeneity for the slope coefficients at the 5% significance level while we have to reject it for the constant terms, even at the 1% level. It is interesting to notice that the equations for which the slope-homogeneity cannot be accepted at the 10% level are those in which non-supported prices are the dependant variables. This result may arise form the relatively low variation in these variables that makes statistical inference less reliable.

Table 4: Chow test results

	F-values of the Chow test		
Dependent variables	Overall homogeneity	Slope homogeneity	Homogeneity in the constant terms
$\ln S_1^*$	0.70	0.57	3.30^c
$\ln PI_{ws,1}$	0.56	0.53	1.36^c
$\ln PI_{ns,1}$	1.14^b	1.11^a	1.18^c
$\ln S_2$	0.91	0.69	5.30^c
$\ln PI_{ws,2}$	0.62	0.62	0.68
$\ln PI_{ns,2}$	1.37^c	1.12^a	3.19^c
$\ln S_3$	0.69	0.61	2.37^c
$\ln PI_{ws,3}$	0.62	0.61	0.98
$\ln PI_{ns,3}$	1.10	1.12^a	0.61

[a]Significant at the 10% level

[b]Significant at the 5% level

[c]Significant at the 1% level

[*] $PI_{ws,i}$:supported price index of brand i,

$PI_{ns,i}$: non-supported price index of brand i,

and S_i: sales of brand i, $i = 1, 2, 3$.

Secondly, to be more confident about our choice of the pooling approach, we estimate RCMs for our applications and compare the Impulse Response Functions (IRFs, they are discussed in the next sub-section) with those of the FEMs. We find that the RCM-IRF results are very similar to those of the FEMs. Thus, based on this

---

[20]We could only apply the pooling tests on 24 stores since for 2 stores there was not enough variation in the data to estimate the unit-by-unit model. In addition, we could not include the exogenous variables either for the same reason.

comparison and on the Chow test results, we feel confident that FEM-based pooling is meaningful for our application.

The fact that our stores are from the same supermarket chain provides a possible explanation for the homogeneity in the slope parameters. The stores attract the same type of people and use marketing instruments in a similar pattern. Hence, the heterogeneity (due to, for example, store or neighborhood size) can be captured by different constants.

7.3 Estimation results

For the identification of the immediate effects we apply the SVAR approach, which is capable of supplementing sample based information with managerial judgement and/or marketing theory (see Dekimpe and Hanssens 2000 and Horváth et al. 2003). We use the following assumptions for the identification. We allow the price variables to have immediate effects on the sales of the brands but do not allow for immediate effects from sales on the prices (feedback requires time). As competitive reaction also takes place with a lag, we do not allow prices to be affected immediately by other brand's prices. We estimate the immediate price elasticities with Full Information Maximum Likelihood. We obtain confidence intervals to determine the 95% confidence interval of the immediate price elasticities and the IRFs that we discuss these later in this section using a bootstrap method[21] based on Benkwitz et al. (1999) with 1000 replications. We show the immediate effects (short-term price elasticities) in Table 5.

Table 5: Price elasticities based on the FEM.

	Price elasticities		
Effects on:	**Brand 1**	**Brand 2**	**Brand 3**
Supported price brand 1	-6.10*	0.49*	1.04*
Non- supported price brand 1	-3.90*	-0.03	0.94*
Supported price brand 2	0.46*	-3.27*	0.78*
Non- supported price brand 2	0.03	-1.97*	0.69*
Supported price brand 3	0.85*	0.33*	-6.89*
Non- supported price brand 3	0.49*	0.35	-4.39*

A * indicates a significant parameter estimate ($\alpha = 0.05$)

These results provide high face-validity of our model since

---

[21]For details on bootstrapping we refer to Benkwitz et al. (1999) or Efron and Tibshirani (1993).

1. All own-brand price elasticities have the right expected sign and are significant.

2. The own-brand price elasticities are much greater (in absolute value) than the corresponding cross-brand price elasticities. This is because the own-brand effects also reflect stockpiling and category expansion effects. Those primary demand effects together tend to account for 50 to 70 percent of the own-brand sales increase due to a temporary price cut for tuna (Van Heerde et al. 2002).

3. All but one of the cross-brand elasticities have the expected sign and most of them are significant. The single cross-brand elasticity with a negative sign is close to zero and is not significant at the 5% significance level.

4. All but one of the supported own- and cross-brand elasticities are larger (in absolute value) than the corresponding non-supported elasticities.

5. Brands with higher own-brand effects usually have higher cross-brand elasticities. Brands that have high own-brand elasticities are generally more sensitive to the effects of prices of competitive brands than brands with lower own-brand elasticities. Brand 3 is the most and brand 2 is the least sensitive in this respect.

Based on the estimated VARX parameters and the identification restrictions that are applied to the immediate relationships among the endogenous variables we can compute the Impulse Response Functions (IRFs). An IRF traces out the effect of an exogenous shock or an innovation in an endogenous variable on all the endogenous variables in the system over time, to provide an answer to the following question: "What is the effect of a shock of size δ in the system at time t on the state of the system at time $t + \tau$, in the absence of other shocks?[22]". In marketing IRFs are often used to estimate the effects of a marketing action on brand performance over time (Dekimpe and Hanssens 1999 , Dekimpe et al. 1999, Horváth et al. 2003, Srinivasan and Bass 2000, and Takada and Bass 1998). In this paper, we consider the effect of price promotions (both supported and non-supported) on own sales. We operationalize the price promotions as one-time deviations from the expected price level. Figure 2 presents for the three different brands how a shock in the own price variables affects their own sales.

Figure 2 shows the dynamic evolution of the sales after a 1% own-price cut. We see that the supported and non-supported price cuts induce high and significant increase in the own sales and we observe some post-promotion dip around the third and/or the fourth week after a supported price promotion. We also see a dip of the IRFs in case of non-supported price promotions, but this dip is not significant at the 5% significance

---

[22]For further details on the IRFs see, for example, Horváth, Kornelis and Leeflang (2002), Hamilton (1994), or Lütkepohl (1993).

Figure 2: Own sales responses to 1% price cuts based on the FEM-VARX model
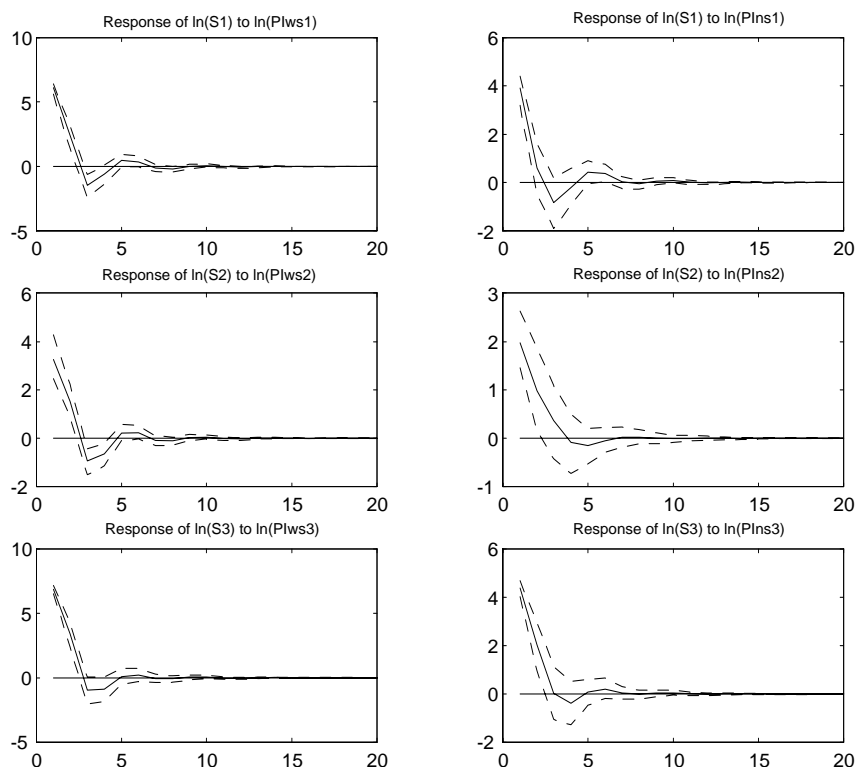
PIwsi refers to supported price index of brand i,
PInsi refers to non-supported price index of brand i,
and Si refers to the sales of brand i, i=1,2,3.

level. For brand 1 the dust-settling period[23] is about $10 - 12$ weeks, both for supported and non-supported own price promotion. For brand 2 and 3 it lasts somewhat shorter, about $5 - 7$ weeks for their supported own price promotion for brand 2, and for about $3 - 4$ for the remainder[24].

The effect of price promotions on competitors' sales (these IRFs are not presented here) are somewhat less prominent. The immediate cross-brand effects are much lower that the corresponding own-brand effects. The dust-settling period for these IRFs lasts

---

[23]The time that is needed for the effects to stabilize (see also Nijs et al. 2001).

[24]Nijs et al. (2001) and Srinivasan et al. (2002) find similar lengths of the dust-settling period (on average).

28

for only about $2 - 4$ weeks.

8 Conclusion, limitations and future research

In this paper we propose a modelling strategy that overcomes one of the main drawbacks of VAR modelling, viz. the degrees-of-freedom problem. This problem occurs frequently in practical VAR applications due the fact that the number of parameters of a VAR model increases quadratically with the dimension of the system of equations. This requires a large number of data points. In addition, unbiasedness of the estimates depends on asymptotics: this puts additional demands on the number of observations. Obviously, the potential danger of this problem lies in obtaining unreliable estimation results due to a shortage of degrees of freedom.

All applications of VAR modelling in marketing to date use specifications on a high aggregation (market-) level. The usual approach for gaining degrees of freedom is to reduce the number of parameters by restricting the model. However, in many cases, the data is also available at a lower aggregation level (e.g. store-level). We propose to gain degrees of freedom not by eliminating parameters, but by increasing the number of observations through pooling data for multiple cross-sectional units.

In this paper, we discuss four different pooling approaches. We start with the two extremes of 'the pooling spectrum'. First we consider the unit-by-unit approach. In this approach the cross-sectional units are not pooled at all; a separate model is estimated for each cross-sectional unit. Secondly, we discuss the Constant Coefficient Model (CCM). This model offers the highest possible degree of pooling, since all parameters in the model are assumed to be constant across all cross-sectional units. The third and the fourth pooling approach are the Fixed Effects Model (FEM) and the Random Coefficients Model (RCM), respectively. These models can be used to strike a balance between the level of pooling and the amount of parameter heterogeneity that is allowed for. For each of the four pooling approaches, we provide details on estimation methods.

We discuss the findings of other authors who show that, in dynamic models, a possible negative side-effect of a high degree of pooling is the introduction of a bias in the parameter estimates, due to not allowing for a sufficient degree of parameter heterogeneity. We illustrate these findings by a small Monte Carlo simulation study in which we compare the results of the pooling approaches and the widely used aggregate approach. The results indicate that the aggregate approach can provide biased and unreliable results. The RCM outperforms the other pooling approaches in terms of efficiency and unbiasedness of the estimated parameters.

We argue that for a particular application, the highest possible degree of pooling and the minimal required amount of cross-sectional heterogeneity depend on the

characteristics of the data. Therefore, we discuss tests that can be used to select the appropriate pooling approach.

We illustrate the proposed modelling strategy by an empirical application. The pooling tests suggest to use the FEM approach. We provide details on the specification and the estimation of this model, and present the empirical results. Both the immediate price elasticities as well as the dynamic effects of price promotions have high face validity. We evaluate the statistical validity of our results using bootstrapped confidence intervals.

The contributions of this paper are that we (1) propose to overcome the degrees-of-freedom problem by pooling (2) provide several pooling approaches that allow for a trade-off between the level of pooling and the degree of heterogeneity (3) compare the efficiency gains of the different pooling approaches by a small simulation study (4) extend existing univariate estimation procedures to a multivariate setting (5) propose to use pooling tests to guide the selection of the appropriate pooling approach (6) use bootstrapped confidence intervals for evaluating the significance of the estimation results.

We leave several areas open for further investigation.

Firstly, in this paper we only investigate the implementation of the 'classical' pooling approaches in a VAR setting. It might be interesting to study alternative approaches. We list several possibilities.

- It might be interesting to discover segments among the cross-sectional units and to estimate a pooled model per segment (using e.g. "fuzzy pooling", see Ramaswamy et al. 1993);

- Another possibility is the Bayesian hierarchical model developed by Western (1998). This approach allows the time-series coefficients to vary across cross sections;

- Bemmaor et al. (1999) suggest to pool the data for some subset of the variables ("partial pooling").

- Bartels (1996) suggests to obtain so-called "fractionally pooled" regression estimators by properly weighting the unit-by-unit estimates based on prior beliefs about the theoretical relevance of disparate observations. He proposes that one should estimate a model allowing for varying degrees of pooling and then take a scientific decision after examining the locus of all such estimates.

Secondly, we assumed a spherical covariance structure for the disturbance term of the VAR models. Hence, we do not allow for spatial interactions of the cross-sectional units. It might be interesting to study how to relax these restrictions.

Thirdly, we only focus on stationary variables. Our approach can be extended to systems with evolving and cointegrated variables.

Fourthly, we do not consider feature, display, and feature and display variables as endogenous in our model. The main reason for this is that they are indicator variables and hence, they require different modelling. Beck, Katz and Tucker (1998) and Beck, Epstein, Jackman and O'Halloran (2001) provide an approach to include a mixture of continuous and indicator variables that might be extended to a VAR setting.

Fifthly, complications may occur in the interpretation of higher order autoregressive processes if we assume a normal distribution of the parameters over the cross sections. Take an AR(2) process for simplicity. Suppose the RC parameter for the AR(1) parameter is 1.2 with standard deviation 0.2, and for the AR(2) parameter it is -0.4 with standard error 0.2. These two RCs capture AR(2) models with parameters 1.0 and 0.0 (the unit roots case), 0.8 and -0.6, (very short cycles), and 1.2 with -0.1 (explosive data) and 1.2 -0.22 (very long cycles). Hence, these RCMs may summarize data that one would not want to summarize in the first place. An approach to overcome this problem would be to use a truncated distribution. This approach may require some Bayesian method (Western 1998). Another solution may be to apply the aforementioned partial pooling approach of Bemmaor et al. (1999), where all parameters are pooled, except the AR parameters.

Sixthly, concerning the bias arising from ignoring heterogeneity in dynamic models, there are several issues that still need to be addressed. The small sample properties of the alternative approaches should be studied more comprehensively through further Monte Carlo studies. This concerns, for example, the behavior of the approaches for (i) data sets with different dimensions, (ii) data with different sources of heterogeneity, and (iii) different VAR models. The small sample behavior of the tests for heterogeneity requires further examination, too.

A The proof that Equation (15) is equivalent with the GLS estimator

Define
$$Y = \left[\mathrm{vec}(Y_1)', \mathrm{vec}(Y_2)', \cdots, \mathrm{vec}(Y_N)'\right]'$$

and
$$Z = \left[\left(Z_1' \otimes I_k\right), \left(Z_2' \otimes I_k\right), ..., \left(Z_N' \otimes I_k\right)\right]'.$$

The GLS estimator of $\mathrm{vec}(H)$ equals:

$$
\begin{aligned}
h &= \left[Z'\Lambda^{-1}Z\right]^{-1} Z'\Lambda^{-1}\mathrm{vec}(Y) \\
&= \left[\sum_{i=1}^{N} (Z_i \otimes I_k)\, \Upsilon_i^{-1}\, (Z_i' \otimes I_k)\right]^{-1} \sum_{i=1}^{N} (Z_i \otimes I_k)\, \Upsilon_i^{-1}\mathrm{vec}(Y_i).
\end{aligned}
$$

Next, we write $\mathrm{vec}(Y_i) = (Z_i \otimes I_k)\, h_i^{OLS}+\mathrm{vec}(\nu_i)$, where ν_i denotes the matrix of fitted residuals, and insert this in the preceding expression:

$$
\begin{aligned}
h &= \left[\sum_{i=1}^{N} (Z_i \otimes I_k)\, \Upsilon_i^{-1}\, (Z_i' \otimes I_k)\right]^{-1} \sum_{i=1}^{N} (Z_i \otimes I_k)\, \Upsilon_i^{-1}\, (Z_i' \otimes I_k)\, h_i^{\mathrm{OLS}} \\
&\quad + \left[\sum_{i=1}^{N} (Z_i \otimes I_k)\, \Upsilon_i^{-1}\, (Z_i' \otimes I_k)\right]^{-1} \sum_{i=1}^{N} (Z_i \otimes I_k)\, \Upsilon_i^{-1}\mathrm{vec}(\nu_i). \quad (18)
\end{aligned}
$$

Subsequently, we recall that $\Upsilon_i = I_T \otimes \Omega_i + \left(Z_i' \otimes I_k\right) \Gamma \left(Z_i \otimes I_k\right)$ and we apply the property: $(A \pm BCB')^{-1} = A^{-1} \mp A^{-1}B\left(C^{-1} \pm B'A^{-1}B\right)^{-1} B'A^{-1}$ to rewrite Υ_i^{-1} in the following way:

$$
\begin{aligned}
\Upsilon_i^{-1} &= \left(I_T \otimes \Omega_i^{-1}\right) - \left[\left(I_T \otimes \Omega_i^{-1}\right) \left(Z_i' \otimes I_k\right) \right. \\
&\quad \left. \times \left(\Gamma^{-1} + \left(Z_i' \otimes I_k\right) \left(I_T \otimes \Omega_i^{-1}\right) \left(Z_i \otimes I_k\right)\right)^{-1} \left(Z_i \otimes I_k\right) \left(I_T \otimes \Omega_i^{-1}\right)\right].
\end{aligned}
$$

The second part of the expression in Equation (18) is equal to zero as $(Z_i \otimes I_k) \left(I_T \otimes \Omega_i^{-1}\right) = \left(Z_i \otimes \Omega_i^{-1}\right) = \left(I_s \otimes \Omega_i^{-1}\right)(Z_i \otimes I_k)$ and because $\nu_i Z_i' = 0$, where $s\,(= Q/k)$ is the number of regressors in one equation. This leaves us with only

the first part:

$$h = \left[\sum_{i=1}^{N} (Z_i \otimes I_k) \, \Upsilon_i^{-1} \left(Z_i' \otimes I_k \right) \right]^{-1} \sum_{i=1}^{N} (Z_i \otimes I_k) \, \Upsilon_i^{-1} \left(Z_i' \otimes I_k \right) h_i^{\mathrm{OLS}}$$

$$= \sum_{i=1}^{N} W_i h_i^{\mathrm{OLS}} \tag{19}$$

$$\text{where } W_i = \left[\sum_{i=1}^{N} (Z_i \otimes I_k) \, \Upsilon_i^{-1} \left(Z_i' \otimes I_k \right) \right]^{-1} (Z_i \otimes I_k) \, \Upsilon_i^{-1} \left(Z_i' \otimes I_k \right).$$

Finally, we use the above result about Υ_i^{-1} and use the same property[25] once more with A, B, and C defined as $A^{-1} = (Z_i \otimes I_k) \left(I_T \otimes \Omega_i^{-1} \right) \left(Z_i' \otimes I_k \right)$, $B = I$ and $C = \Gamma$. The desired result is then obtained by noting that

$$
\begin{aligned}
(Z_i \otimes I_k) \, \Upsilon_i^{-1} \left(Z_i' \otimes I_k \right) &= (Z_i \otimes I_k) \left(I_T \otimes \Omega_i^{-1} \right) \left(Z_i' \otimes I_k \right) \\
&\quad - \Big[(Z_i \otimes I_k) \left(I_T \otimes \Omega_i^{-1} \right) \left(Z_i' \otimes I_k \right) \\
&\qquad \times \left\{ \Gamma^{-1} + (Z_i \otimes I_k) \left(I_T \otimes \Omega_i^{-1} \right) \left(Z_i' \otimes I_k \right) \right\}^{-1} \\
&\qquad \times \quad (Z_i \otimes I_k) \left(I_T \otimes \Omega_i^{-1} \right) \left(Z_i' \otimes I_k \right) \Big] \\
&= \left((Z_i Z_i' \otimes \Omega_i^{-1})^{-1} + \Gamma \right)^{-1} \\
&= (V_i + \Gamma)^{-1}.
\end{aligned}
$$

$\square$

---

[25]Note that we could only use this property in the system of equations context because of the special structure of an (unrestricted) VARX model, the panel homoscedasticity and because of the block-diagonality of the Λ matrix.

References

Arellano, M. and Bond, S. (1991), 'Some tests of specification for panel data: Monte carlo evidence and an application to employement equations', *Review of Economic Studies* **58**, 277–297.

Baltagi, B. H., Hidalgo, J. and Li, Q. (1996), 'A nonparametric test for poolability using panel data', *Journal of Econometrics* **75**, 345–367.

Bartels, L. M. (1996), 'Pooling disparate observations', *American Journal of Political Science* **40**, 905–942.

Beck, N. (2001), 'Time-series-cross-section data', *Statistica Neerlandica* **55**, 111–133.

Beck, N., Epstein, D., Jackman, S. and O'Halloran, S. (2001), Alternative models of dynamics in binary time-series-cross-section models: The example of state failure. Working Paper, University of California.

Beck, N. and Katz, J. N. (1995), 'What to do (and not to do) with time-series-cross-section data', *American Political Science Review* **89**, 634–647.

Beck, N. and Katz, J. N. (2001), Random coefficient models for time-series-cross-section data: The 2001 version. Working Paper, University of California.

Beck, N., Katz, J. N. and Tucker, R. (1998), 'Taking time seriously: Time-serious-cross-section analysis with a binary dependent variable', *American Journal of Political Science* **42**, 1260–1288.

Bemmaor, A. C., Franses, P. H. and Kippers, J. (1999), 'Estimating the impact of display and order merchandising support on retail brand sales: Partial pooling with examples', *Marketing Letters* **10**, 87–100.

Benkwitz, A., Lütkepohl, H. and Wolters, J. (1999), Comparison of bootstrap confidence intervals for impulse responses of german monetary systems. CEPR Discussion Paper No. 2208.

Binder, M., Hsiao, C. and Pesaran, M. H. (2002), Estimation and inference in short panel vector autoregressions with unit roots and cointegration. Working Paper, University of Maryland.

Bun, M. J. G. (2001*a*), Accurate Statistical Analysis in Dynamic Panel Data Models, PhD thesis, Tinbergen Institute.

Bun, M. J. G. (2001*b*), Testing poolability in a system of regressions with nonspherical disturbances. Working Paper, Tinbergen Institute.

Canova, F. and Ciccarelli, M. (2000), Forecasting and turning point predictions in a bayesian panel VAR model. Universitat Pompeu Fabra, Economics and Business Working Paper: 443.

Chow, G. (1960), 'Tests of equality between sets of coefficients in two linear regressions', *Econometrica* **28**, 591–605.

Dekimpe, M. G. and Hanssens, D. M. (1995), 'The persistence of marketing effects on sales', *Marketing Science* **14**, 1–21.

Dekimpe, M. G. and Hanssens, D. M. (1999), 'Sustained spending and persistent response: A new look at long-term marketing profitability', *Journal of Marketing Research* **36**, 397–412.

Dekimpe, M. G. and Hanssens, D. M. (2000), 'Time-series models in marketing: Past, present and future', *International Journal of Research in Marketing* **17**, 183–193.

Dekimpe, M. G., Hanssens, D. M. and Silva-Risso, J. M. (1999), 'Long-run effects of price promotions in scanner markets', *Journal of Econometrics* **89**, 269–291.

Efron, B. and Tibshirani, R. J. (1993), *An Introduction to Bootstrap*, Chapman and Hall, New York.

Enders, W. (1995), *Applied Econometric Time Series*, John Wiley and Sons, INC., New York.

Franses, P. H. (1998), *Time Series Models for Business and Economic Forecasting*, Cambridge University Press, Cambridge.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Hanssens, D. M. (1980), 'Market response, competitive behavior, and time series analysis', *Journal of Marketing Research* **27**, 470–485.

Haque, N. U., Pesaran, M. H. and Sharma, S. (2000), *Neglected Heterogeneity and Dynamics in Cross-Country Savings Regressions*, Elsevier Science B.V., chapter 3, pp. 53–82.

Holtz-Eakin, D., Newey, W. and Rosen, H. S. (1988), 'Estimating vector autoregressions with panel data', *Econometrica* **56**, 1371–1395.

Holtz-Eakin, D., Newey, W. and Rosen, H. S. (1989), 'The revenues-expenditures nexus: Evidence from local government data', *International Economic Review* **30**, 415–429.

Horváth, C., Kornelis, M. and Leeflang, P. S. H. (2002), What marketing scholars should know about time series analysis. SOM Research Report No. 02F17, University of Groningen.

Horváth, C., Leeflang, P. S. H., Wieringa, J. E. and Wittink, D. R. (2003), Dynamic analysis of a marketing system based on aggregated and pooled store data. Paper under review.

Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge University Press, Cambridge.

Hsiao, C. and Sun, B. (2000), *To Pool or Not to Pool Panel Data*, North-Holland, chapter 9, pp. 181–198.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. and Lee, T.-C. (1985), *The Theory and Practice of Econometrics*, second edn, John Wiley and Sons, New York.

Kornelis, M. (2002), Modeling Market Shake-Ups Using Time Series Data, PhD thesis, SOM, University of Groningen.

Leeflang, P. S. H. and Wittink, D. R. (2000), 'Building models for marketing decisions: Past, present and future', *International Journal of Research in Marketing* **17**, 105–126.

Leeflang, P. S. H., Wittink, D. R., Wedel, M. and Naert, P. A. (2000), *Building Models for Marketing Decisions*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Lehr, C. S. (1999), 'Banking on fewer children: Financial intermediation, fertility and economic development', *Journal of Population Economics* **12**, 567–590.

Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Magnus, J. R. and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley and Sons, Chichester.

Moriarty, M. and Salamon, G. (1980), 'Estimation and forecast performance of a multivariate time series model of sales', *Journal of marketing research* **17**, 558.

Neyman, J. and Scott, E. (1948), 'Consistent estimates based on partially consistent observations', *Econometrica* **16**, 1–32.

Nijs, V. R., Dekimpe, M. G., Steenkamp, J.-B. E. M. and Hanssens, D. M. (2001), 'The category- demand effects of price promotions', *Marketing Science* **20**, 1–22.

Parsons, L. J. and Schultz, R. L. (1976), *Marketing Models and Econometric Research*, North Holland Publishing Company, Amsterdam.

Pesaran, M. H. and Smith, R. (1995), 'Estimating long-run relationships from dynamic heterogenous panels', *Journal of Econometrics* **68**, 79–113.

Ramaswamy, V., Desarbo, W. S., Reibstein, D. J. and Robinson, W. T. (1993), 'An empirical pooling approach for estimating marketing mix elasticities with PIMS data', *Marketing Science* **12**, 103–122.

Robertson, D. and Symons, J. (1992), 'Some strange properties of panel data estimators', *Journal of Applied Econometrics* **7**, 175–189.

Rousseau, P. L. and Wachtel, P. (2000), 'Equity markets and growth: Cross-country evidence on timing and outcomes, 1980-1995', *Journal of Banking and Finance* **24**, 1933–1957.

Sims, C. A. (1980), 'Macroeconomics and reality', *Econometrica* **48**, 1–48.

Srinivasan, S. and Bass, F. M. (2001), Diagnosing competitive responsiveness: Disentangling retailer-induced and manufacturer-induced actions. Paper presented at the MSI Conference on Competitive Responsiveness, Boston.

Srinivasan, S., Leszczyc, P. T. L. P. and Bass, F. M. (2000), 'Market share response and competitive interaction: The impact of temporary, evolving and structural changes in prices', *International Journal of Research in Marketing* **17**, 281–305.

Srinivasan, S., Pauwels, K., Hanssens, D. M. and Dekimpe, M. G. (2002), Do promotions benefit manufacturers, retailers or both? Working Paper, Erasmus Research Institute of Management, Erasmus University of Rotterdam.

Swamy, P. A. V. B. (1971), *Statistical Inference in Random Coefficient Regression Models*, Springer-Verlag, New York.

Takada, H. and Bass, F. M. (1998), 'Multiple time series analysis of competitive marketing behavior', *Journal of Business Research* **43**, 97–107.

Van Heerde, H. J., Leeflang, P. S. H. and Wittink, D. R. (2000), 'The estimation of pre- and postpromotion dips with store- level scanner data', *Journal of Marketing Research* **37**, 383–395.

Van Heerde, H. J., Leeflang, P. S. H. and Wittink, D. R. (2001), 'Semiparametric analysis to estimate the deal effect curve', *Journal of Marketing Research* **38**, 197–215.

Van Heerde, H. J., Leeflang, P. S. H. and Wittink, D. R. (2002), Flexible decompositions of price promotion effects using store-level scanner data. MSI Working Paper Series, No. 02-107.

Western, B. (1998), 'Causal heterogeneity in comparative research: A bayesian hierarchical modelling approach', *Americal Journal of Political Science* **42**, 1233–1259.

Zellner, A. (1962), 'An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias', *Journal of American Statistical Association* **57**, 348–368.