Pooling Multi-country Data: Short Data and Multi-generations of Technologies

Towhidul Islam
Department of Marketing and Consumer Studies
University of Guelph, Canada

Nigel Meade
Tanaka Business School
Imperial College, London, UK

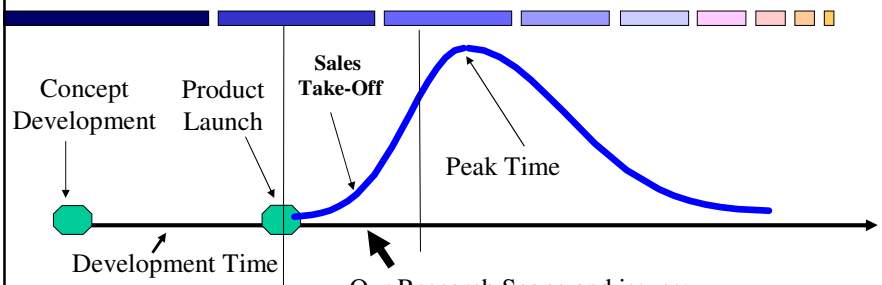
1

Pooling Multi-country Data: Short Data and Multi-generations of Technologies

- **Our Research Objectives**
- **Pooling short data:**
  - **Single generation multi-country data**
  - **Multi-generation, Multi-country data**
    - » **Multi-generation Diffusion Model**
    - » **Duration Times between Key events – Hazard Model Approach**
- **Further Work**

2

Research Objectives



We need forecasts:
- Survey
- Conjoint
- Discrete Choice

Our Research Scope and issues:
- Short Data
- Dramatic increase in sales (say 400%)

Requires resources for:
- Manufacturing
- Inventory
- Distribution
- Sales staff

3

• Research Stream: Pooling Single generation multi-country data

- **Islam, T., Fiebig, D. and Meade, N. (2002), "Modelling Multinational Telecommunications Demand with Limited Data", International Journal of Forecasting, 18, 605-624.**
- **Islam, T. and Fiebig, D. (2001), "Modelling the Developments of Supply-Restricted Telecommunications Markets", Journal of Forecasting, 20, 249-264.**

Discussion on : Islam, Fiebig and Meade (2002)

- **Data Description**
- **Innovation Diffusion Model**
- **Pooling Techniques**
- **Summary results**

4

Data and Model: Pooling Single generation multi-country data

- Data: Digital Cellular- 16 countries, ISDN – 16 countries and Facsimile – 38 countries
  - Short data: On average 5 -6 observations per country
- Innovation Diffusion Model: Linearised Gompertz

$$y_{ct} = f(m_c, \phi) + \varepsilon_{ct}$$
$$y_{ct} = \ln(Y_{ct} / Y_{ct-1}) = \phi[\ln m_c - \ln Y_{ct-1}] + \varepsilon_{ct}$$
$$\ln(Y_{ct} / Y_{ct-1}) = \phi \ln m_c - \phi \ln Y_{ct-1} + \varepsilon_{ct}$$

Intercept **Slope**

5

Pooling Techniques: Pooling Single generation multi-country data

- **Fixed Effect Model**
  - » Market saturation, m_c for each country, slope or growth coeff. (i.e. ϕ) is common across all the countries
- **Cross-Sectionally Varying (CSV) Model**
  - » Market saturation, m_c for each country depends of a number of covariates , slope or growth coeff. (i.e. ϕ) is common across all the countries
  - » Covariates: GDP, Connection Charge, Call Charge, Market Size at Introduction
- Random Effect Model
- Random Coefficient Model

6

- Plausibility of coefficient Estimates: **Pooling Single generation multi-country data**

Market Saturation: Digital Cellular

Covariates	Estimates	P-Value
Constant	1.220	0
GDP	0.078	0
Connection Charge	-0.076	0
Call Charge	-0.003	0.826
Market size	0.008	0.003

Growth: 0.096 (p =0)

Market Saturation: Fax Connections

Covariates	Estimates	P-Value
Constant	1.079	0
GDP	0.166	0
Waiting time	-0.033	0
Market size	0.007	0.826

Growth: 0.138 (p =0)

7

Plausibility of Market Saturation Estimates: **Pooling Single generation multi-country data**

1998

	Main Tel/100	Cell/100	Fixed Effect	CSV	
Austria	46.9	22.8	17.5	44.1	
Belgium	46.5	17.3	4.5	24.7	
Denmark	61.8	33.9	19.8	91.1	
Finland	54.9	51.1	11.5	106.6	
France	56.4	19.0	5.8	26.2	
Germany	53.8	16.6	5.1	11.9	
Italy	44.0	29.7	0.5	84.1	
Norway	55.5	41.5	32.5	72.9	
Sweden	68.2	41.1	24.7	76.0	
Switzerland	64.0	23.3	2.4	32.3	
UK	56.8	23.2	9.5	34.6	

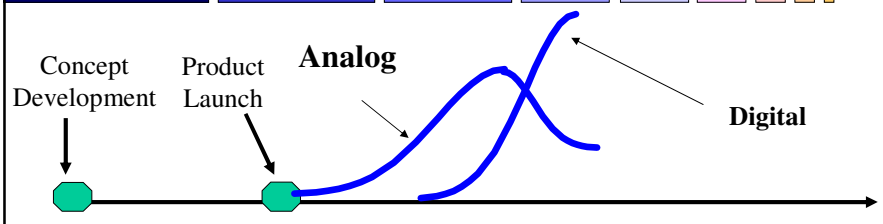
8

Model Selection and Forecasting Performance: Pooling Single generation multi-country data

- Forecasting Performance of 4 pooling methods and a deterministic trend model
  - » MAPE
  - » Example Digital Cellular – 1 to 16 step ahead forecasts
  - » CSV model better upto 3 years
  - » Fixed effect is better for 4 years – no need to forecast covariates
- Model Selection
  - » AIC and BIC
  - » The most preferred model is CSV, then Fixed Effect

9

Research Stream: Multi-generation Diffusion Model



- **Switching Adoption bet. Technologies: Multi-generation**
- **Short Data**
- **Investigate Possibility of Pooling Multi-country Data?**
- **Covariate Effects?**
- **Forecasting Performance?**

10

Research Stream: Multi-generation Diffusion Model

- **Work in Progress**
  - **Sample of Earlier Work**
  - **Modeling Approach**
  - **Data and Covariates**
  - **Some Preliminary Results**

11

Sample of Earlier Work: Multi-generation Diffusion Model

- **Multi-generation: Single Country**
  - **Norton and Bass (1987), Management Science**
  - **Mahajan and Muller (1996), TFSC**
  - **Danaher, Hardie and Putsis (2001), JMR**
  - **Pae and Lehman (2003), JAMS**
  - **Islam and Meade (1997),TFSC**
- **Single generation: Multi-country**
  - **Ganesh, Kumer, and Balasubramanian (1997), *JAMS***
  - **Talukder, Sudhir and Ainsle (2002), Marketing Science**
  - **Islam, Fiebig, and Meade (2002), *IJF***

Extend our earlier work

12

Data and Covariates: Multi-generation Diffusion Model

- **Multi-generation Cellular Data: Analog and digital**
  - **Complete data set for about 70 countries**
- **Covariates**
  - **Covariates that influence the speed of innovation can be classified into 3 major types.**
    - » **Characteristics of Innovation**
    - » **Country Characteristics**
    - » **Context.**
- **We have collected about 50 Covariates (both Time-Varying and Time-Invariant)**

13

Covariates: **Multi-generation Diffusion Model**

- Innovation Characteristics
- Wealth Related
- Access and Availability of Information
- Demographics
- Culture
- Heterogeneity
- Geographical
- Political

14

Modeling Approach: Norton and Bass Model

- **$\text{Analog}_{ct} = f(m_{1c}, p_a, q_a) + \varepsilon_{1ct}$**
- **$\text{Digital}_{ct} = f(m_{1c}, m_{2c}, p_d, q_d) + \varepsilon_{2ct}$**
  - where $\varepsilon_{1ct} \sim N(0, \sigma^2_{1c})$ and $\varepsilon_{2ct} \sim N(0, \sigma^2_{2c})$
- **m_{1c} = Market Saturation for Analog**
- **m_{2c} = Incremental Market Saturation for Digital**
- **p_a and p_d = Coefficient of innovation ~ captures the early growth of the technology**
- **q_a and q_d = Coefficient of imitation (word of mouth effect) ~ captures the long term growth of the technology**

15

Multi-generation Diffusion Model

Analog yearly connections

$$y_{1ct} = m_{1c}\left(\left(F_{1ct} - F_{1ct-1}\right) - \left(F_{1ct}F_{2c(t-\tau)} - F_{1ct-1}F_{2c(t-\tau-1)}\right)\right) + \varepsilon_{1ct}$$

$$\varepsilon_{1ct} \sim N\left(0, \sigma^2_{1ct}\right) and\ \sigma^2_{1ct} = \alpha_{1c0} + \alpha_{1c1}m_{1c}$$

$$y_{2ct} = m_{2c}\left(F_{2c(t-\tau)} - F_{2c(t-\tau-1)}\right) + m_{1c}\left(F_{1ct}F_{2c(t-\tau)} - F_{1ct-1}F_{2c(t-\tau-1)}\right) + \varepsilon_{2ct}$$

$$\varepsilon_{2ct} \sim N\left(0, \sigma^2_{2ct}\right) and\ \sigma^2_{2ct} = \alpha_{2c0} + \alpha_{2c1}m_{1c} + \alpha_{2c2}m_{2c}$$

Digital Cellular yearly connections

16

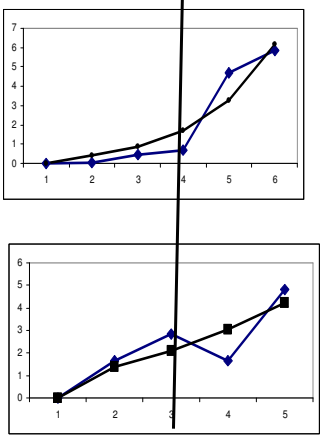
Modeling Approach: Norton and Bass Model

- **The Objective is to maximize the log likelihood function**

$$\sum_{\text{Model(M)}} \sum_{\text{Generation(G)}} \sum_{\text{Country(C)}} \sum_{\text{Time(T)}} \ln\left(f\left(\frac{\varepsilon_{MGCT}}{\sigma^2_{MGC}} \right) \right)$$

17

Sample Pooling of Two Countries: Multi-generation Diffusion Model



Digital: Fit and Forecasts

	Parameters	Estimates	Prob.
Common	Pa	0.008	0.00
	qa	1.045	0.00
	Pd	0.003	0.17
	qd	0.621	0.00
Slovania	m1	0.453	0.00
	m2	145.415	0.02
	Var(analog)	0.011	0.00
	Var (digital)	0.751	0.68
Slovak	m1	0.394	0.25
	m2	201.973	0.03
	Var(analog)	0.020	0.00
	Var (digital)	0.909	0.90

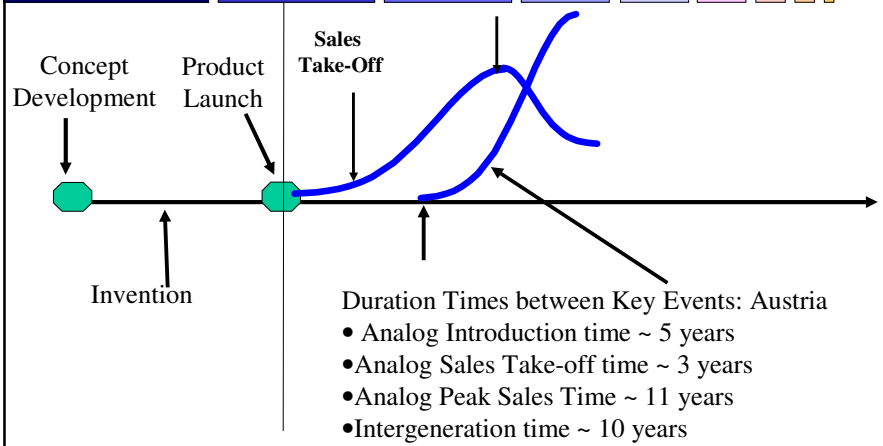
18

Further Work

- **Further Work**
  - **Pooled Estimates: First Regional and then International**
  - **Incorporate Covariates**
  - **Forecasting Performance**

19

Research Stream: Duration Time between Key events – Hazard Model Approach



Duration Times between Key Events: Austria
- Analog Introduction time ~ 5 years
- Analog Sales Take-off time ~ 3 years
- Analog Peak Sales Time ~ 11 years
- Intergeneration time ~ 10 years
- Digital Introduction time ~ 3 years
- Digital Sales Take-off time ~ 2 years
- Digital Peak Sales Time ~ Censored

20

Duration Time between Key events – Hazard Model Approach

- Data and Covariates – 80 countries, analog and digital cellular connections
- Research objectives
  - Predict duration probabilities of future events conditional on covariates and past events
- Modeling Approach
  - Duration times ~ modeling using different hazard models e.g. Weibul, Erlang, Logistic with Gamma or Gaussian Heterogeneity
  - Dependence Structure among the Key Events
    - Multivariate Copulas – We have used this approach in our earlier works
      - » Meade and Islam (2003) "Modeling Dependence between the Times to International Adoption of Two Related Technologies", TFSC
      - » **Islam and Meade (2004) Modelling the evolution of inter-purchase times for consumer packaged products**

21

Further Investigation

- We also wish to address the issue of sample "Selectivity"
  - All together we are collecting data from 150 countries but at the end we will end up analyzing only 70 Countries.
  - As our selection of 70 countries are not random, ignoring "selectivity" will lead to biased estimates
  - We shall not be able to generalize the results

22