E-Companion for

"*A Generalized Norton-Bass Model for Multigeneration Diffusion*"

Zhengrui Jiang
College of Business, Iowa State University
2340 Gerdin Business Building
Ames, IA 50011-1350
zjiang@iastate.edu

Dipak C. Jain
INSEAD
Boulevard de Constance
77305 Fontainebleau, France
dipak.c.jain@insead.edu

I. GNB Model for the Three-Generation Scenario

The expressions for $y_2(t)$ and $Y_2(t)$ shown in Equations (11) and (12) are valid only if G2 is the latest generation available. Now suppose a generation 3 (G3) is introduced at τ_3 ($\tau_3 \geq \tau_2$). Similar to the shifting from G1 to G2 after τ_2, some of the existing and potential adopters of G2 will consider adopting G3 after τ_3. In order to derive the number of leapfrogging and switching adoptions between time τ_3 and t, we first define

$$\tilde{y}_2(t) = [m_2 + m_1 F_1(t)] f_2(t - \tau_2) + m_1 f_1(t) F_2(t - \tau_2), t \geq \tau_2, \text{ and}$$

$$\tilde{Y}_2(t) = [m_2 + m_1 F_1(t)] F_2(t - \tau_2), t \geq \tau_2.$$

Here $\tilde{y}_2(t)$ and $\tilde{Y}_2(t)$ represent the adoption rate and the cumulative number of adoptions of G2, assuming that G3 is never introduced. After the introduction of G3, however, a portion of those who otherwise would have adopted G2 will substitute it with G3 instead. Similar to leapfrogging from G1 to G2, we assume that the proportion of leapfrogging from G2 and G3 is a direct result of the diffusion of adoption concerning G3. Hence, the rate of leapfrogging from G2 to G3 at time t equals

$$u_3(t) = \tilde{y}_2(t) F_3(t - \tau_3), \; t \geq \tau_3.$$

In addition to leapfrogging, switching from G2 to G3 also starts to occur after time τ_3. Analogous to that between G1 and G2, the rate of switching from G2 to G3 is

$$w_3(t) = \tilde{Y}_2(t) f_3(t - \tau_3), \; t \geq \tau_3.$$

Based on $u_3(t)$ and $w_3(t)$ and their cumulative forms $U_3(t)$ and $W_3(t)$, we obtain the non-cumulative and cumulative rate of adoptions for G2 and G3:

$$y_2(t) = \begin{cases} \tilde{y}_2(t), & \tau_2 \leq t < \tau_3, \\ \tilde{y}_2(t) - u_3(t), & t \geq \tau_3, \end{cases}$$

$$Y_2(t) = \begin{cases} \tilde{Y}_2(t), & \tau_2 \leq t < \tau_3, \\ \tilde{Y}_2(t) - U_3(t), & t \geq \tau_3. \end{cases}$$

$$y_3(t) = m_3 f_3(t-\tau_3) + u_3(t) + w_3(t),\ t \geq \tau_3,$$

$$Y_3(t) = m_3 F_3(t-\tau_3) + U_3(t) + W_3(t),\ t \geq \tau_3,$$

where m_3 represents the market potential unique to G3. Similar to the two-generation scenario, we also derive the number of units-in-use for each generation:

$$S_1(t) = Y_1(t) - W_2(t),$$

$$S_2(t) = Y_2(t) - W_3(t),$$

$$S_3(t) = Y_3(t).$$

It can be verified that these results are consistent with the NB model with three generations.

II. Comparison of Models that Differentiate Leapfrogging and Switching

	Mahajan & Muller (1996)	**Jun & Park (1999) Type I Model**	**Danaher et al. (2001)**	**Jiang (2010)**	**GNB**
Data Suitable for Parameter Estimation	units-in-use	units-in-use	units-in-use	sales	units-in-use OR sales
Closed-Form Expressions as Direct Functions of Model Parameters and Time	No	No*	No*	sales	units-in-use AND sales
Consistency with the Norton-Bass Model	No	No	No	No	Yes
Number of Generations Derived	four-generation	N-generation	two-generation	two-generation	N-generation
Continuous/Discrete Time	continuous	discrete	discrete	continuous	continuous
Leapfrogging Multiplier	constant	time-varying	time-varying	constant	time-varying
Switching Multiplier	constant	time-varying	time-varying	time-varying	time-varying
Marketing-Mix Variables	No	Yes	Yes	No	Yes

* Jun and Park (1999) and Danaher et al. (2001) also provide expressions for units-in-use. However, in order to obtain the number of units-in-use for period *t*, the numbers for all preceding periods must be calculated first. Using the GNB model, on the other hand, we can directly calculate the number for any time period.

III. Evaluating the Rates of Leapfrogging and Switching

The separation of leapfrogging and switching adoptions is a critical aspect of the GNB model. Due to the limitation of the available data, however, we are not able to directly test the leapfrogging and switching terms of the GNB model in our empirical analyses. In fact, the same is true for other existing models that also differentiate leapfrogging from switching. Lacking an adequate real dataset, we resort to simulation to generate appropriate datasets, which are then used to evaluate the GNB model's performance regarding the separation of leapfrogging and switching adoptions.

We base our simulation on the adoptions and substitutions of analog and digital cellular services. In order to simulate a customer's adoption behavior, we need to make assumptions regarding the probability of leapfrogging and switching at any given time. Since none of the existing multigeneration diffusion models have been empirically tested for leapfrogging and switching, we use three models, i.e., Danaher et al.'s model, Jun and Park's Type I model, and the GNB model for this purpose.[1] These models are selected because they all deliver good fit to our empirical data, as shown previously. We assume that each customer's leapfrogging or switching behavior is equally likely governed by one of the three models. We refer to this simulation method as the *benchmark model*, which is used to evaluate how well a particular model separates leapfrogging and switching adoptions.

To be consistent with real-world settings, the simulation is based on the parameters estimated from the US cellular subscriptions data. We simulate a total of 1 million potential customers, of which 35.2% are potential adopters of the analog service (counted in m_1), and 64.5% are the potential adopters added after the introduction of the digital service (counted in m_2). This ratio between m_1 and m_2 represents the average ratio between the two parameters estimated based on the three models. Given the market size for both generations, we first simulate the potential customers' adoption behaviors based on the benchmark model, and then repeat the simulation three more times, each time assuming that all potential customers' adoption behaviors follow only one of three models. For each model, the adopted parameter values are the same as those estimated from the US cellular subscriptions data, as summarized in Table E1.

Table E1. Parameter Values Used in Simulation

GNB		**Danaher et al. (2001)**		**Jun & Park (1999) Type I Model**	
Para.	Value	Para.	Value	Para.	Value
p	0.00389	p	0.00176	c^*	5.354
q_1	0.431	q_1	0.471	q_1	0.153
q_2	0.554	q_2	0.835	q_2	0.363

* c is the constant parameter in Jun and Park's model.

For ease of comparison, we display the key simulation results in two figures. The rates of leapfrogging associated with different models are compared in Figure E1. The rates of switching are compared in Figure E2. From Figure E1, we observe that the rates of leapfrogging vary widely across the different models. Importantly, among the three models, the curve representing the GNB model seems to be the closest to the curve associated with the benchmark model. To confirm this, we directly compare the

[1] Jun and Park's Type I model does not explicitly express the rate of leapfrogging. Consistent with our definition of leapfrogging and how it is calculated by other models, we assume that leapfroggers in their model are the portion of adopters who would have adopted G1 if G2 was not available, but choose to adopt G2 once it is introduced.

three competing models' rates of leapfrogging against that of the benchmark model; the mean absolute derivations (MADs) are summarized in the first row of Table E2. We can see that the GNB model leads to the smallest MAD. Danaher et al.'s model and Jun and Park's model result in MADs that are 108% and 67%, respectively, higher than that of the GNB model. These results suggest that the GNB model is the best option in projecting the rate of leapfrogging between the two generations.

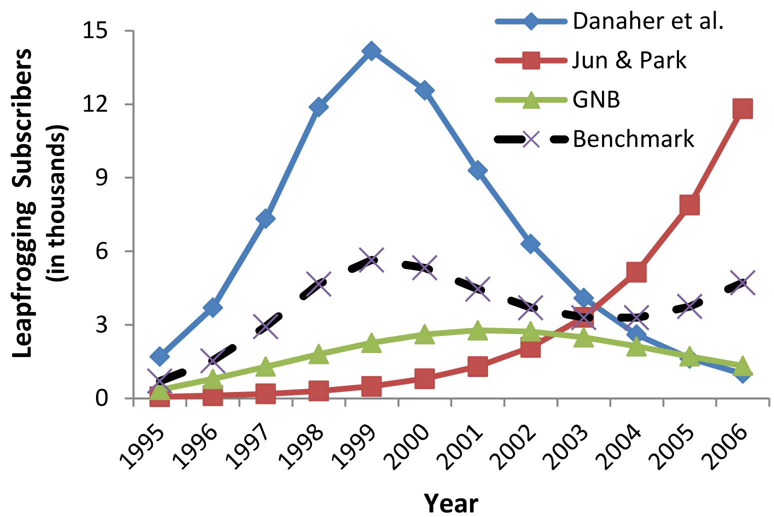

Figure E1. Comparison of the Rates of Leapfrogging Projected by Different Models

Table E2: Projections of Leapfrogging and Switching: Mean Absolute Deviation from the Benchmark Model

	Danaher et al. (2001)	Jun & Park (1999)	GNB
Leapfrogging	3,776	3,063	1,814
Switching	7,343	8,636	3,477

Similarly, we find from both Figure E2 and the MADs shown in the second row of Table E2 that the rate of switching projected by the GNB model is the closest to the rate of the benchmark model. This shows that among the three compared models, the GNB model is the safest option in estimating the rate of switching between the two generations.

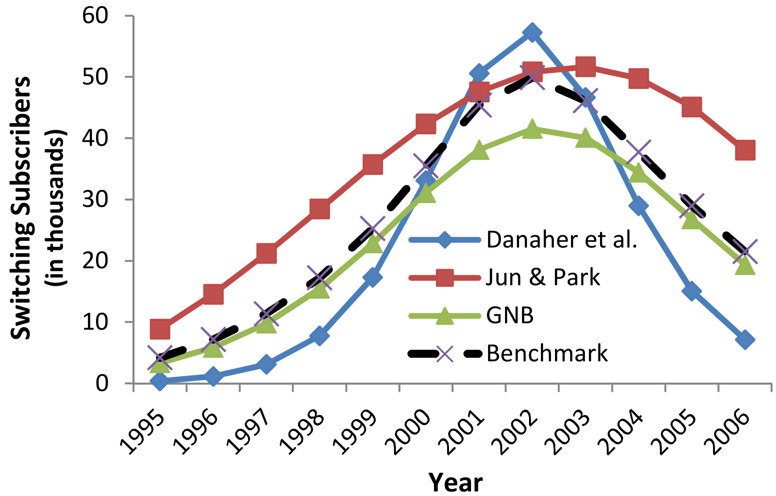

Figure E2. Comparison of the Rates of Switching Projected by Different Models

The evaluation based on the distance to the benchmark model essentially assumes that all three models have an equal chance of being the "true" model, i.e., the accurate representation of customers' adoption behaviors in the real-world. What if all customers' adoptions are governed by only one model? Next, we use a cross-validation method to further evaluate how each model performs in separating leapfrogging from switching adoptions. Specifically, we assume each time that a particular model is the "true" model, and then check how the other two models perform compared to this "true" model. Since the GNB model is guaranteed to perform the best when it is assumed to be the "true" model, we report only the comparisons when Danaher et al's and Jun and Park's models are treated as the "true" model. The results are summarized in Tables E3 and E4. From Table E3, we conclude that if Danaher et al.'s model reflects the customers' true adoption behaviors, the GNB model produces better estimations of leapfrogging and switching adoptions than does Jun and Park's Type I model. Specifically, for leapfrogging, the MAD associated with Jun and Park's model is 50% higher than that of the GNB model; for switching, the MAD of Jun and Park's model almost doubles that of the GNB model. If Jun and Park's Type I model is the "true" model, Table E4 shows that the GNB model provides estimations closer to the assumed "true" model than does Danaher et al.'s model. The MADs associated with Danaher et al.'s models are 175% (for leapfrogging) and 27% (for switching) higher than those associated with the GNB model.

Table E3: Mean Absolute Deviation when Danaher et al.'s Model is "True"

Jun & Park Type I Model		**GNB**	
Leapfrogging	Switching	Leapfrogging	Switching
6,837	15,371	4,570	7,817

Table E4: Mean Absolute Deviation when Jun & Park's Type I Model is "True"

Danaher et al.'s Model		GNB	
Leapfrogging	Switching	Leapfrogging	Switching
6,837	15,371	2,484	12,113

In summary, based on the analyses reported in this section, we conclude that the GNB model compares favorably against competing models in the estimation of the rates of leapfrogging and switching.

IV. SAS Codes for the GNB Model

a. Fitting the Three-Generation DRAM Sales Data

```
data DRAM;
input t g1 g2 g3;
datalines;
…          …          …          …
;

proc print data= DRAM; run;

proc model  CONVERGE=.0000001 MAXITER = 10000 method=marquardt;
endogenous g1 g2 g3;
exogenous t;
        parameters
                p1 =  0.01
                p2 =  0.01
                p3 =  0.01
                q1 = 0.3
                q2 = 0.3
                q3 = 0.3
                m1 = 3000000
                m2 = 1500000
                m3= 2000000;
        p2=p1; p3=p1; /* equal p */
        con1=(p1+q1)*(p1+q1)/p1;
        con2=(p2+q2)*(p2+q2)/p2;
        con3=(p3+q3)*(p3+q3)/p3;
        e1=exp(-(p1+q1)*t);                de1=1+(q1/p1)*e1;
        e2=exp(-(p2+q2)*(t-12));           de2=1+(q2/p2)*e2;
        e3=exp(-(p3+q3)*(t-29));           de3=1+(q3/p3)*e3;

if t<13 then /* second generation introduced in period 13 */
    do;
        g1=m1*con1*e1/de1/de1;
        g2=0;
        g3=0;
    end;
else if t<30 then /* third generation introduced in period 30 */
    do;
        g1=m1*con1*e1/de1/de1*(1-(1-e2)/de2);
        g2=(m2+m1*(1-e1)/de1)*con2*e2/de2/de2 + m1*con1*e1/de1/de1*(1-e2)/de2;
        g3=0;
```

```
     end;
else
     do;
          g1=m1*con1*e1/de1/de1*(1-(1-e2)/de2);
          g2= ( (m2+m1*(1-e1)/de1)*con2*e2/de2/de2 + m1*con1*e1/de1/de1*(1-e2)/de2 ) * (1-(1-e3)/de3);
          g3= (m3+ ( m2+m1*(1-e1)/de1 )*(1-e2)/de2 ) * con3*e3/de3/de3
            + ( (m2+m1*(1-e1)/de1)*con2*e2/de2/de2 + m1*con1*e1/de1/de1*(1-e2)/de2 ) * (1-e3)/de3;
     end;

fit  /fiml;
run;
```

b. Fitting the Two-Generation Cellular Subscription Data

```
data USCell;
input t g1 g2;
datalines;
…         …         …
;

proc print data= USCell; run;

proc model  CONVERGE=.0000001 MAXITER = 10000 method=marquardt;
endogenous g1 g2;
exogenous t;
          parameters
                    p1        = 0.01
                    p2        = 0.01
                    q1        = 0.4
                    q2        = 0.4
                    m1        = 50000000
                    m2        = 200000000;
          p1=p2; /* equal p */
          e1=exp(-(p1+q1)*t);         de1=1+(q1/p1)*e1;
          e2=exp(-(p2+q2)*(t-11));  de2=1+(q2/p2)*e2;

if t<12 then /* second generation introduced in period 12 */
     do;
          g1=m1*(1-e1)/de1;
          g2=0;
     end;
else
     do;
          g1=m1*(1-e1)/de1*(1-(1-e2)/de2);
          g2=(m2+m1*(1-e1)/de1)*(1-e2)/de2;
     end;

fit  /fiml;
run;
```