Retail Sales Generation: A Methodological Comparison of Econometric Estimation and Calibration

Howard Hao-Chun Chuang

Mays Business School, Texas A&M University, College Station, TX 77843, USA

hchuang@mays.tamu.edu

Rogelio Oliva

Mays Business School, Texas A&M University, College Station, TX 77843, USA

roliva@tamu.edu

Olga Perdikaki

Mays Business School, Texas A&M University, College Station, TX 77843, USA

operdikaki@mays.tamu.edu

Abstract

We take a production economics perspective to formulate a sales generating function that incorporates traffic and labor in retail stores. The estimation of this function is the basis for the development of a full system dynamics model that will enable managers to take a holistic view on labor planning process. To recover the structural parameters, we compare regression-based estimation and optimization-based calibration. This comparative approach allows us to assess the usefulness and applicability of new econometric estimation paradigms in the context of system dynamics modeling. We conclude by briefly describing possible extensions to our modeling efforts, expected outcomes, and methodological implications in terms of calibration.

Key words: system dynamics; econometrics; calibration; production function; retailing.

1. Introduction

It is widely acknowledged that effective management of labor is critical to successful retail operations (Fisher et al., 2007). Staffing in the retail sector is complex because store associates fulfill service-related tasks (e.g., check-out, returns, shopping assistance) and production-like tasks (i.e., in-store logistics) (Ton, 2009). Volatile store traffic further complicates the process of determining staffing levels. Apparently there are intricate relationships between labor, traffic, and sales performance. However, limited studies have been conducted to better understand how labor and traffic interact with each other and further affect store sales (Lam et al., 1998).

To fill in the gap, Mani et al. (2011) and Perdikaki et al. (2012) use a data set of high-end women's retail apparel stores to empirically assess the emerging issue of labor-traffic mismatch. Complementary to their investigation, in this paper we develop a sales generating function grounded on production economics and calibrate the function using the same data set, which contains hourly observations of *traffic*, *sales*, and *labor hours* over a whole year. The estimation of this function is the basis for the development of a full system dynamics model that will enable managers to take a holistic view on labor planning process.

To calibrate the proposed formulation, we adopt various econometric models to estimate the function and identify algebraic relationships to recover structural parameters. We show that optimization-based calibration and regression-based estimation have comparable performance. More importantly, the sales generating function has fairly strong explanatory power of sales variation. This comparative approach allows us to assess the usefulness and applicability of new econometric estimation paradigms in the context of system dynamics modeling. We conclude by briefly describing possible extensions to our modeling efforts, expected outcomes, and methodological implications in terms of calibration.

2. A sales generating function in retailing

Grounded on neoclassical economic theory, previous studies (e.g., Lam et al., 1998, Mani et al., 2011) capture the generation of sales by employing a production function to operationalize how traffic and labor are converted into store sales. The transcendental production function used in those studies generalizes the popular Cobb-Douglas production function without assuming constant elasticity of substitution (Griffin et al 1987). In this study we first specify a generalized power production function (Janvry 1972) that is highly flexible and subsumes the transcendental production function. The sales-generating function is specified as:

$$\mathrm{S_t} = \mathrm{N_t}^{\alpha} e^{\gamma\left(\phi_t/\phi^*\right)} \tag{1}$$

where t: day index, $\mathrm{S_t}$: Sales, $\mathrm{N_t}$: Traffic, $\phi_t = \mathrm{L_t}/\mathrm{N_t}$ and $\mathrm{L_t}$: Labor.

In line with recent empirical findings (Perdikaki et al., 2012), we posit that the labor-to-traffic ratio ϕ_t is a key element of sales generation, that is, what matters is not the labor available per se, but how labor compares to the store traffic. While ϕ_t will normally take small values—in many store formats customers can find merchandize without the support of a sales representative—formulating the ratio as $\mathrm{L_t}/\mathrm{N_t}$ makes it an increasing function of labor. The effectiveness of ϕ_t will depend on how much it deviates from the reference point ϕ^*, which defines the base labor effects and results in $1 sales per customer when $\phi_t = \phi^*$. The formulation ϕ_t/ϕ^* constitutes *labor adequacy* that drives retail store sales because it affects both service and conformance quality (Oliva and Sterman, 2001). Nonetheless, Sterman (2000, Ch. 14) proposes an alternative formulation that is more robust and easier to interpret. The formulation is intuitive in that $\phi_t=\phi^*$ leads to the natural reference line (i.e., $1 sales per customer), unlike the exponential function in which the basis has to depend on γ even when $\phi_t=\phi^*$. Accordingly, we revise the generalized power function to:

$$S_t = N_t^{\alpha}\left(\frac{\phi_t}{\phi^*}\right)^{\gamma} \tag{2}$$

The above function is well-behaved and has the desired properties. The right panel of figure 1 illustrates the response of sales to N_t^{α}. The parameter α is the elasticity of traffic. Since an increase in traffic should only increase sales at a diminishing rate (Mani et al., 2011), we expect $0<\alpha\leq 1$, where $\alpha=1$ leads to constant return-to-scale. The second factor of equation (2) denotes the effect of labor-to-traffic ratio on sales. The left panel of figure 1 shows how the effect of labor on sales varies with ϕ_t. A higher γ implies that the transaction value is more responsive to changes in ϕ_t, and the reference point ϕ^* is the point at which sales per customer transaction is 1.

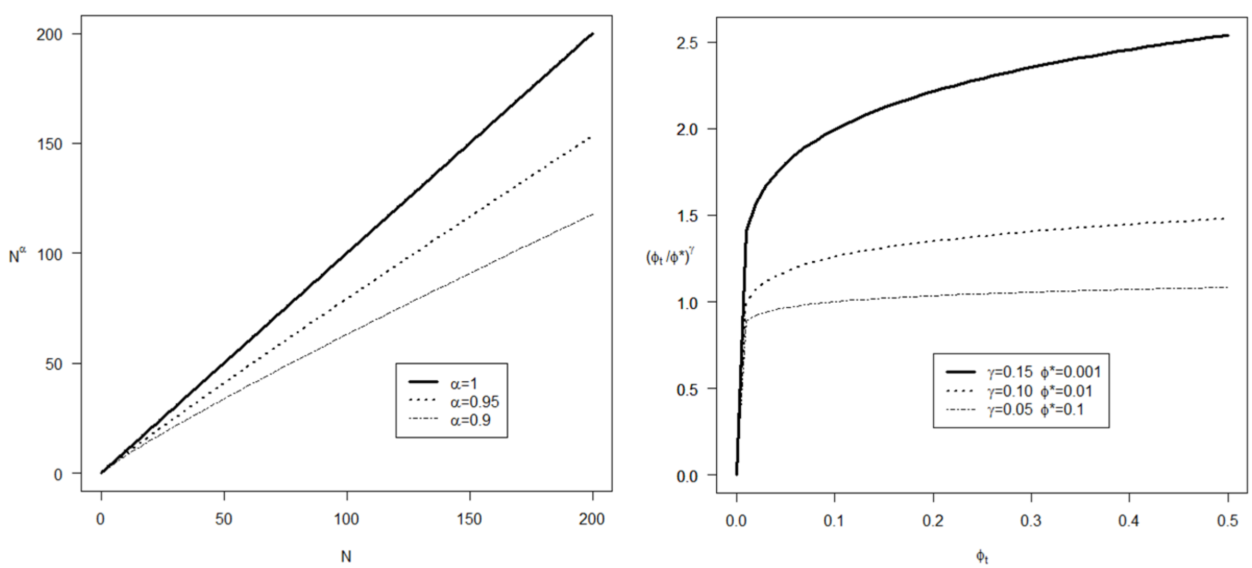

Figure 1: Input response of proposed sales generating function

Substituting ϕ_t by its constituting observable elements, we can rewrite the (2) as:

$$S_t = N_t^{\alpha}\left(\frac{L_t}{N_t\phi^*}\right)^{\gamma}$$

After taking the natural log and after some algebra, the function becomes:

$$\mathrm{Ln}(S_t) = (\alpha-\gamma)\mathrm{Ln}(N_t) + \gamma\mathrm{Ln}(L_t) - \gamma\mathrm{Ln}(\phi^*) \tag{3}$$

which makes sales an linear function of traffic and labor that can be estimated through regression.

$$\mathrm{Ln}(\mathrm{S_t}) = \mathrm{b_0} + \mathrm{b_1}\mathrm{Ln}(\mathrm{N_t}) + \mathrm{b_2}\mathrm{Ln}(\mathrm{L_t}) + \varepsilon_\mathrm{t} \tag{4}$$

Although the reference point ϕ^* is not observable, since ϕ^* is intrinsic to the store and time-invariant, in principle it should be orthogonal to the noise ε_t. Note that the random noise ε_t is not part of the structural sales generating function (2) and requires distributional assumptions to facilitate empirical estimation. Thus, we can employ estimated coefficients to recover structural parameters α, γ, and ϕ^*. Using (3) and (4) we obtain the following relationships:

$$\alpha = \mathrm{b_1} + \mathrm{b_2}; \gamma = \mathrm{b_2}; \phi^* = \exp(-\frac{\mathrm{b_0}}{\mathrm{b_2}}) \tag{5}$$

It might be worth noting that, since we are only estimating a single store, the b_0 parameter is a mixture of the store fixed effects and the reference point ϕ^*. Once we introduce more stores into the estimation and fully utilize the panel data set, we should be able to separate these effects and obtain a more accurate estimate of ϕ^*, which is intrinsic to each store.

In the next section we derive parameter estimates using regression-based approaches and model calibration (Oliva, 2003). The estimates are then used to generate simulated outcomes for further comparison.

3. Estimation Results

Figure 2 shows the sales generating function and data. The left panel represents equation (2) that we directly calibrate to recover those structural parameters. The system dynamics representation of (2) can be viewed as a micro structure of a full model so that the calibration exercise is essentially a partial model simulation (Oliva, 2003).

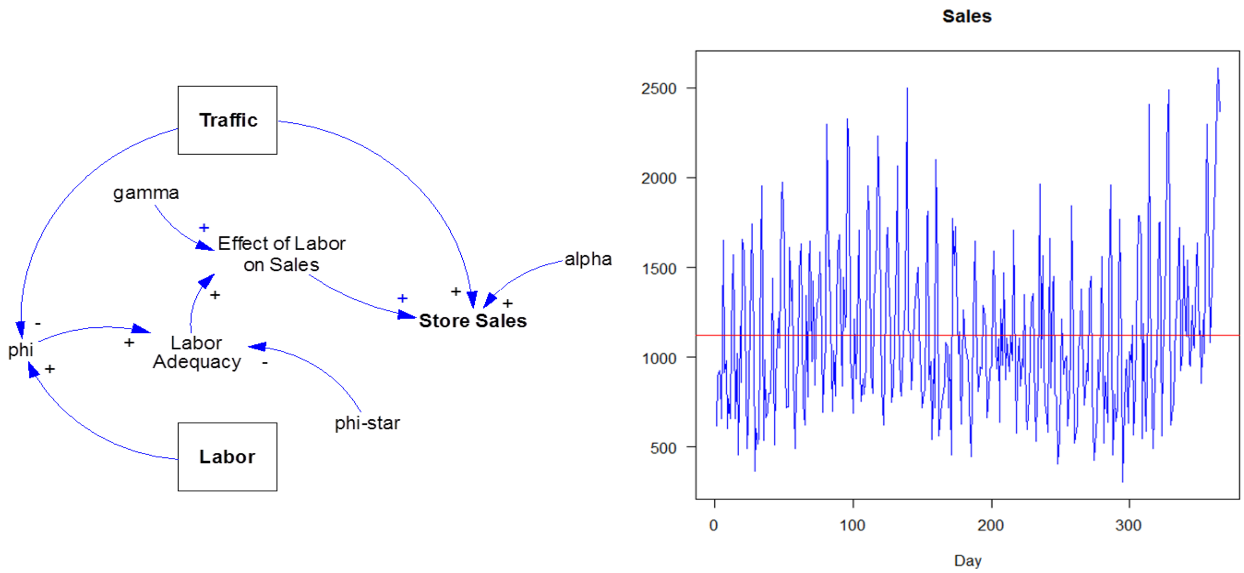

Figure 2: Sales generating function and data

The right panel of figure 2 illustrates the average hourly sales on each day of a store we selected for testing. The horizontal line is the average sales level around $1120. Different stores in our data set in general exhibit similar sales pattern. There is strong evidence of a weekly cycle with peaks during the weekend. To tackle the noisy variation of sales, we estimate (4) using various regression techniques. Afterwards we obtain α, γ, and ϕ^* based on the algebraic relationships in (5).

Table 1 shows parameter estimates and their corresponding standard errors (in parentheses). Because the estimated ϕ^* is a very small number, we take the natural log and report it in the exponential form. The first row illustrates ordinary least square (OLS) estimates with White robust standard error (Wooldridge, 2001). Although we found no significant collinearity, there is strong evidence of serial correlation in residuals that violate OLS assumptions. We resolved the issue using Newey-West standard error (see the second row of Table 1), which accounts for heteroskedasticity and autocorrelation up to certain lags (lags = 7 in our case). The corrected OLS estimates, however, may still be biased due to the potential endogeneity between labor and sales. The simultaneity bias (Wooldridge, 2001) could occur in that the labor drives sales and sales affects staffing decisions. We tackled the endogeneity issue using the instrumental variable (IV) approach (Wooldridge, 2009). For the time series regression models, we adopted lagged labor (with lag = 7 days given the presence of strong weekly cycles) as an instrument.

Table 1: Parameter estimates

	n	b_0	b_1	b_2	α	γ	ϕ^*
OLS	361	3.430 (0.167)	0.801 (0.033)	0.150 (0.064)	0.951 (0.066)	0.150 (0.064)	$e^{-22.876}$
OLS New-West	361	3.430 (0.162)	0.801 (0.031)	0.150 (0.061)	0.951 (0.066)	0.150 (0.061)	$e^{-22.876}$
IV (lag7_llabor)	351	3.366 (0.297)	0.813 (0.027)	0.163 (0.142)	0.976 (0.150)	0.163 (0.142)	$e^{-20.650}$
Calibration	361				0.989 (0.003)†	0.186 (0.001)†	$e^{-18.159}$

† Inferred from the MLE 95% confidence interval reported by Vensim assuming confidence intervals at ±1.96*SE.

Note that the commonly used two-stage least square (2SLS) estimation does not account for heteroskedasticity and serial correlations in our data. Instead, we used an efficient and robust generalized method of moments (GMM) estimator (Baum et al., 2007) to obtain results reported in the third row of Table 1. Although the parameter estimates were slightly adjusted when we used the instrumental variable, in line with Perdikaki et al. (2012), the test of endogeneity did not reveal significant biases, and the difference in estimates are not statistically significant. Finally, in the fourth row of Table 1, we report calibrated estimates obtained from non-linear optimization algorithms in Vensim (Ventana Systems, 2010).

The estimates derived from regression and calibration are not identical and we are interested in exploring the performance of different econometric techniques to system dynamics modeling. With actual sales, the purpose is to see whether regression estimates provide a better/worse fit to historical data. Table 2 illustrates the summary of different metrics of error (simulated sales – actual sales) and Theil inequality statistics (Sterman, 1984).

Table 2: Summary statistics of historical fit

Statistic	Simulation Calibrated Est.	Simulation IV Est.	Simulation OLS Est.
R-square	0.678	0.677	0.676
Mean Abs. Percent Error	0.199	0.194	0.194
Mean Square Error	62911.262	63424.487	64087.700
Root Mean Square Error	250.821	251.842	253.156
Theil's Bias	0.000	0.006	0.013
Theil's Variation	0.099	0.106	0.128
Theil's Covariation	0.901	0.888	0.859

In spite of its slightly higher mean absolute percent error, calibration-based simulation outperforms IV- and OLS-based simulation in nearly all dimensions listed in Table 1. However, the large mean square errors and Thiel statistics imply that we need to expand the generating function to better explain sales dynamics. Interestingly, unlike Morecroft (1977) who showed that OLS outperformed 2SLS-IV in simulation experiments, we found that robust GMM IV estimates result in slightly better simulated outcomes than OLS estimates. Even though we detected no significant evidence of endogeneity empirically, conceptually the endogeneity is not uncommon in estimating production function estimation with observational data (Varian, 1991). The concern of endogeneity supports why system dynamics calibration may preferred to over regression-based approaches to estimate the sales generating process since system dynamics is by design to address bi-directional interactions of causes and effects. Therefore, the concern of simultaneity (Wooldrige, 2001) that makes IV often preferred over OLS in empirical economics is not indeed a validity threat to system dynamics modeling.

Although simple, this preliminary function captures a fair amount of sales variation in the absence of feedback loops and additional controls. We are aware that the proposed sales generating function has not fully captured salient features of store operations, and we are currently working on a more comprehensive model to better approximate the data generating process and explain the macro-behavior of store operations. Natural extensions will be the full use the panel data that we have available (45 different stores) as well as for controlling for different fixed store characteristics and time variant factors due to seasonality.

Our work aims to shed light on the negative impact of myopic workforce allocation. Our model tackles understaffing, which is driven by minimizing payroll expenses and becomes so common in retailing. Seeing the empirical evidence that increasing the amount of labor at a store enhances conformance quality and indirectly improves profitability (Ton, 2009), we hope to identify staffing policies that are economically favorable and improve store execution. Salmon (1989) argues that retailing has shifted into the age of *execution* and the quality of execution will distinguish winners from losers. We posit that store labor is key to execution, which "makes what should happen in retail stores actually happen" (Fisher, 2004). Our analysis hopes to derive some useful insights for organizational design in the retail sector.

Although the current results are preliminary in terms of the development of the full model, we believe our findings from estimation and calibration trigger some interesting methodological insights that we discuss in the closing section.

4. Methodological implications

In this paper we compared optimization-based calibration with regression-based estimation using observational data. We took a structural approach promoted by empirical economists who are in favor of calibration (e.g., Kydland and Prescott, 1996; Dawkins et al., 2001). We first developed the intuition for what the generating function of sales ought to be, and we brought econometric rigor into the process of estimating model parameters. The preliminary finding that OLS and IV did not provide superior performance has important implications because system dynamics has been accused of a lack of statistical foundation by economic modelers since its infancy (Barlas, 1996). In response to those critics, system dynamicists have attempted to improve model validity by incorporating various statistical techniques (Sterman, 1984; Dogan, 2007). Our perspective is slightly different from the traditional thought and we want to show that calibration actually has a strong empirical foundation as pointed out by leading econometricians (Hansen and Heckman, 1996; Dawkins et al., 2001).

The system dynamics community should recognize that calibration is more than a testing strategy and potentially can make system dynamics an eligible tool to answer empirically interesting questions in positive economics. Moreover, system dynamics is able to specify the social process that accounts for how causes bring about their effects, unlike regression analysis that does not entail on any particular data generating mechanism (Morgan and Winship, 2007). It is the ability to mimic the data generating process that enables calibration and simulation to address what-is, what-might-be, and what-should-be that cannot be evaluated under an ordinary regression modeling framework (Kydland and Prescott, 1996; Burton and Obel, 2011). We simply want to highlight the fact that calibration is more than a validity test and it is a reasonable strategy to empirically derive key parameters of the micro-structure of the model (Oliva, 2003).

Once the model is fully articulated, we also intend to explore the possibility of conducting a two-step calibration. While a well-developed system dynamics model has high face and internal validity, high external validity is not easy to achieve in general. Hansen and Heckman (1996) discuss a two-step scheme in which only part of the empirical data is used to infer structural parameters in the first step called ‘calibration’. In the second step called ‘verification’ the model is simulated and the results are compared to the hold-out part of the data to better evaluate external validity. We have time-series data of different operational metrics on a daily basis throughout a whole year. So, we may be qualified to perform the two-step procedure without losing too much information in step-one calibration.

References

Barlas, Y. 1996. Formal aspects of model validity and validation in system dynamics. *System Dynamics Review* 12(3): 183-210.

Baum. C. F., Schaffer, M. E. and Stillman, S. 2007. Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *STATA Journal* 7(4): 465-506.

Burton R. M. and Obel, B. 2011. Computational modeling for what-is, what-might-be, and what-should-be studies—and triangulation. *Organization Science* 22(5): 1195-1202.

Dawkins, C., T. N. Srinivasan, J. Whalley. 2001. Calibration. In Heckman, J.J., Leamer, E. (eds.), *Handbook of Econometrics*. 5: Chapter 58. Elsevier Science.

Dogan, G. 2007. Bootstrapping for confidence interval estimation and hypothesis testing. *System Dynamics Review* 23(4): 415-436.

Fisher, M. 2004. To me it's a store. To you it's a factory. *ECR Journal* 4(2): 9-18.

Fisher, M.L., Krishnan, J. and Netessine, S. 2007. Retail store execution: an empirical study. *Working Paper*. The Wharton School, University of Pennsylvania, Philadelphia.

Griffin, R. C., J. M. Montgemery, M. E. Rister. 1987. Selecting functional form in production function analysis. *Western Journal of Agricultural Economics* 12(2): 216-227.

Hansen, L. P. and Heckman, J. J. 1996. The empirical foundation of calibration. *Journal of Economic Perspectives* 10(1): 87-104.

Janvry, A. D. 1972. The generalized power production function. *American Journal of Agricultural Economics* 54(2): 234-237.

Kydland, F. E. and Prescott, E. C. 1996. The computational experiment: An econometric tool. *Journal of Economic Perspectives* 10(1): 69-85.

Lam, S., M. Vandenbosch, M. Pearce. 1998. Retail sales force scheduling based on store traffic forecasting. *Journal of Retailing* 74(1): 61-88.

Mani, V., S. Kesavan, J. M. Swaminathan. 2011. *Understaffing in retail stores: Drivers and consequences*. Working Paper, Smeal College of Business, Pennsylvania State University, State College, PA.

Morecroft, J. D. W. 1977. A comparison of ordinary least squares and instrumental variables estimation of a dynamic growth model. System Dynamics Group Working Paper D-2729, Sloan School of Management, MIT, Cambridge, MA.

Morgan, S. L., C. Winship. 2007. *Counterfactuals and causal inference: Methods and Principles for Social Research*. Cambridge University Press, New York.

Oliva, R., J. D. Sterman. 2001. Cutting corners and working overtime: Quality erosion in the service industry. *Management Science* 47(7): 894-914.

Oliva, R. 2003. Model calibration as a testing strategy for system dynamics models. *European*

Journal of Operational Research 151(3): 552-568.

Perdikaki, O., S. Kesavan, J. M. Swaminathan. 2012. Effect of traffic on sales and conversion rates of retail stores. *Manufacturing & Service Operations Management* 14(1): 145-162.

Salmon, W.J. 1989. Retailing in the age of execution, *Journal of Retailing* 65(3): 368-378.

Sterman, J.D. 1984. Appropriate summary statistics for evaluating the historical fit of system dynamics models. *Dynamica* 10 (Winter), 51–66.

Steman, J. D. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill, Boston.

Ton, Z. 2009. The effect of labor on profitability: The role of quality. *Working Paper*. Harvard Business School, Boston.

Varian, H. R. 1991. *Microeconomic Analysis*. 3rd ed. Norton, New York.

Ventana Systems, Inc. 2010. *Vensim DSS Software 5.11a*, Ventana Systems Inc., Harvard, MA.

Wooldridge, J. M. 2001. *Analysis of Cross Section and Panel Data*. The MIT Press, Boston.

Wooldridge, J. M. 2009. *Introductory Econometrics: A Modern Approach*. 4th ed. South-Western College Publishing, Cincinnati, OH.