

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Estimating high-dimensional demand systems in the presence of many binding non-negativity constraints

Daniel L. Millimet^{a,*}, Rusty Tchernis^b^a Southern Methodist University and IZA, United States^b Indiana University and NBER, United States

ARTICLE INFO

Article history:

Available online 16 September 2008

JEL classification:

C11
C31
D12
L66

Keywords:

Demand systems
Translog
Coherency
Data augmentation
Markov Chain Monte Carlo
Gibbs' Sampling

ABSTRACT

Two econometric issues arise in the structural estimation of consumer or producer demand systems in the presence of many binding non-negativity constraints. Firstly, most existing methods entail the evaluation of multivariate probability integrals. Secondly, the issue of statistical coherency must be addressed. We circumvent both of these issues using Gibbs' Sampling, along with data augmentation and rejection sampling. We illustrate our method using several simulated data sets.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Welfare analysis of various government interventions that alter the relative price of goods, such as taxes or subsidies, as well as the impact of non-governmental actions that alter relative prices, such as natural disasters or shocks that disrupt supplies of imported goods, requires knowledge of the behavioral response of economic agents to price changes. Acquisition of this knowledge requires estimation of complete systems of demand equations, typically at a very disaggregate level. However, once one turns to disaggregated consumption data, one quickly confronts the issue of zero consumption of some goods by at least some agents. In other contexts (e.g., when the allocation of goods is rationed or when estimating models of labor supply), one may confront corner solutions or kinks in the budget set at values other than zero. Unfortunately, the proper estimation of complete systems of consumer or input demand equations from survey data containing highly disaggregated data with a substantial number of corner solutions – at zero or elsewhere – is an ongoing issue in applied demand analysis.

* Corresponding address: Department of Economics, Box 0496, Southern Methodist University, Dallas, TX 75275-0496, United States. Tel.: +1 214 768 3269; fax: +1 214 768 1821.

E-mail address: millimet@mail.smu.edu (D.L. Millimet).

Derivation of an econometric model from the maximization of a random direct utility or profit function subject to the Kuhn–Tucker conditions characterizes the primal solution to the consumer or firm problem (Wales and Woodland, 1983). This method, however, rules out the use of more flexible demand specifications for which no explicit specification of the direct utility or profit function can be given. The dual solution consists of deriving demand systems from indirect cost or utility functions – including popular flexible functional forms such as the translog – by specifying virtual (or reservation) prices which are dual to the Kuhn–Tucker conditions (Lee and Pitt, 1986, 1987). While the efficiency and theoretical attractiveness of the Lee and Pitt (1986, 1987) approach is well acknowledged, its usefulness in practice is severely constrained by two econometric realities. The first is the curse of dimensionality. The unobserved nature of the virtual prices for non-consumed goods leads to a censoring problem, requiring the evaluation of multivariate probability integrals when evaluating the likelihood function. The second relates to the internal consistency of the model (Ransom, 1987; van Soest and Kooreman, 1990; van Soest et al., 1993). In the demand system literature, this is referred to as the (statistical) coherency problem (Gourieroux et al., 1980). In light of these two difficulties, Arndt (1999, p. 207) concludes that while the dual approach is theoretically attractive and efficient, it is “extremely difficult or impossible to apply in practice”.

In response to the issue of dimensionality, the computational burden can be reduced by placing a factor structure on the random

components of the indirect utility function (Lee and Pitt, 1986) or by using simulated maximum likelihood methods (Kao et al., 2001). However, the coherency issue has been much more difficult to address. For the model to be well-behaved, there must exist a one-to-one mapping between all feasible realizations of the stochastic components of the model and all possible vectors of demands. Unfortunately, coherency is not guaranteed over the entire parameter space of the model. In such cases where the parameters of a model stray outside the coherent region, the sum of the probabilities for all demand regimes is not unity and maximum likelihood estimates are inconsistent (van Soest et al., 1993; Arndt, 1999). Interestingly, the problem of statistical incoherency arises when “the implications of economic theory, particularly the curvature conditions, are violated for some sets of values of parameters, exogenous variables, and error terms” (Arndt, 1999, p. 210).

To prevent the model from being incoherent, researchers have utilized a number of strategies when estimating demand systems with binding non-negativity constraints. Firstly, one may impose the parametric restrictions required to guarantee global concavity of the translog or other flexible functional forms (e.g. Chakir et al., 2004). However, since concavity of the cost or utility function is a sufficient, but not necessary condition, for coherency (van Soest and Kooreman, 1990), this approach is unduly restrictive, thereby destroying the very flexibility these functional forms seek to provide (Lau, 1978; Jorgenson and Fraumani, 1981; Gallant and Golub, 1984; Diewert and Wales, 1987; Terrell, 1996). Secondly, one may utilize globally concave functional forms. However, such functional forms impose unrealistic restrictions on behavior, particularly with disaggregate demands and micro-level data. Hausman (1985) and MaCurdy et al. (1990) encounter similar problems in models of individual labor supply, verifying that the restrictions needed to guarantee global concavity impose undesirable constraints on income and wage elasticities. Thirdly, a number of empirical researchers have opted to simply check *ex post* if concavity holds at each data point and report it as a ‘statistic’ reflecting the reasonableness of the restrictions imposed by demand theory. Since the parameter estimates are inconsistent if concavity is rejected at even a single data point, such *ex post* methods are clearly unsatisfactory.

Finally, two other alternatives have been used in the literature. Firstly, the Amemiya–Tobin approach has been applied, where estimation is via full information maximum likelihood, quasi-maximum likelihood, or various two-step estimators. Secondly, maximum entropy (ME) estimation has been used. The Amemiya–Tobin approach departs from the fully structural approach of Lee and Pitt (1986, 1987) by, firstly, deriving deterministic budget shares and appending additive error terms and, secondly, ignoring the role of virtual prices (e.g. Dong et al. (2004)). While the reduced form approach circumvents the coherency issue (Yen et al., 2003), ignoring the role of virtual prices leads to inconsistent estimates when, in fact, reservation prices matter (Arndt, 1999). The ME approach has been applied in Arndt (1999) and Golan et al. (2001). While ME permits the use of virtual prices for non-consumed goods, the impact of coherency on the properties of the ME estimator is not understood (Arndt, 1999), nor are the asymptotic properties of the ME estimator in nonlinear applications such as in the case of censored demand systems (Arndt, 1999; Meyerhoefer et al., 2005).

Given the drawbacks of the solutions preferred to date, the development of estimation methods that are fully consistent with the underlying structural model, avoid the need for high-dimensional integration, and impose only the minimum set of restrictions – those that ensure the model is locally coherent – is needed. In this paper, we aim to fill this gap by building on several recent, promising additions to this literature. Specifically,

Terrell (1996) and O’Donnell et al. (1999) employ Bayesian techniques to impose the theoretical concavity conditions locally in demand systems containing only interior solutions, and Pitt and Millimet (1999) utilize Gibbs sampling techniques, along with rejection sampling, to impose local concavity in demand systems containing many binding non-negativity constraints (see also Barnett et al. (1991)). In particular, we extend the work of Lee and Pitt (1986, 1987) by using the Gibbs sampling technique, as in Terrell (1996), along with the data augmentation algorithm (Tanner and Wong, 1987), as in Pitt and Millimet (1999), to solve both the dimensionality and the coherency problem. Data augmentation eliminates the need for high-dimensional integration, and rejection sampling permits the straightforward imposition of only those conditions necessary for coherency – *coherency at each data point* – rather than global regularity.

Prior to continuing, it is worth noting that this paper builds heavily on the unpublished study by Pitt and Millimet (1999). The approach put forth here differs, though, in three important respects. Firstly, Pitt and Millimet (1999) utilize the Gibbs Sampler in a purely classical framework. Here, priors are included in the algorithm. Secondly, the sampling methods utilized herein are refined and even eliminated in some places. Not only is the algorithm presented here more convenient computationally, but the extensive use of cumbersome rejection sampling in Pitt and Millimet (1999) may have interfered with the convergence properties of the Gibbs Sampler. Finally, whereas Pitt and Millimet (1999) utilize computationally more taxing non-linear SUR estimation within each iteration of the Gibbs Sampler when estimating consumer demand systems, here we utilize a more convenient approach based on an auxiliary regression.

The remainder of the paper is organized as follows: Section 2 presents the dual approach to the agent’s maximization problem when there are binding non-negativity constraints; Section 3 discusses estimation of the model; Section 4 presents some simulation results; and, Section 5 concludes.

2. Model

2.1. Dual approach to binding non-negativity constraints

Following Lee and Pitt (1986), let $H(v; \theta, \varepsilon)$ be an indirect utility function defined as

$$H(v; \theta, \varepsilon) = \max_q \{U(q; \theta, \varepsilon) | vq = 1\} \quad (1)$$

where the utility function, $U(\cdot; \theta, \varepsilon)$, is strictly quasi-concave, continuous, and monotonic, q is a K -dimensional vector of quantities, v is a similarly sized vector of normalized (by income) market prices, θ is a vector of unknown parameters, and ε is a vector of random components. The *latent* demand equations $q(v; \theta, \varepsilon)$ for the set of K goods are derived by applying Roy’s Identity:

$$q_k = \frac{\frac{\partial H(v; \theta, \varepsilon)}{\partial v_k}}{\sum_{j=1}^K v_j \frac{\partial H(v; \theta, \varepsilon)}{\partial v_j}}, \quad k = 1, \dots, K. \quad (2)$$

These demand equations are latent because they are not required to be non-negative in (1). Let x_k represent the non-negative observed demands. The latent demands q_k correspond to x_k as follows: there exists a vector of positive virtual prices π which can exactly support these zero demands (or any other allocation) as long as the preference function is strictly quasi-concave, continuous, and monotonic (Neary and Roberts, 1980).

The derivation of the virtual price functions for the non-consumed goods is manageable since the denominator in (2)

drops out of the function. Without loss of generality, let the observed demand for the first K_0 goods be zero. The virtual prices $\pi_k(v_{K_0+1}, \dots, v_K)$ are solved from the equations

$$\frac{\partial H(\pi_1(\bar{v}), \dots, \pi_{K_0}(\bar{v}), \bar{v}; \theta, \varepsilon)}{\partial v_k} = 0, \quad k = 1, \dots, K_0 \quad (3)$$

where $\pi_k(\bar{v})$ is the virtual price of good k and \bar{v} is the vector of market prices of the positively consumed goods, indexed by $k = K_0 + 1, \dots, K$. The market prices \bar{v} are also the virtual prices for the consumed goods as they exactly support the observed positive demands. The remaining (positive) demands are given by

$$x_k = \frac{\frac{\partial H(\pi_1(\bar{v}), \dots, \pi_{K_0}(\bar{v}), \bar{v}; \theta, \varepsilon)}{\partial v_k}}{\sum_{j=1}^K v_j \frac{\partial H(\pi_1(\bar{v}), \dots, \pi_{K_0}(\bar{v}), \bar{v}; \theta, \varepsilon)}{\partial v_j}}, \quad k = K_0 + 1, \dots, K. \quad (4)$$

The Eqs. (4) are estimable and the parameters of the latent demand equations (2) can be identified by estimating this conditional demand system. Selection among different regimes, where a regime is defined by the set of consumed goods at the optimum, is done by a comparison of the virtual and market prices. The regime in which the first K_0 goods are not consumed is characterized by the conditions

$$\pi_k(\bar{v}) \leq v_k, \quad k = 1, \dots, K_0. \quad (5)$$

This characterization follows directly from the Kuhn–Tucker conditions.

2.2. Translog demand system

To implement the preceding methodology, we utilize the translog indirect utility function of Christensen et al. (1975), given by

$$H(v; \theta, \varepsilon) = - \sum_{k=1}^K (\alpha_k + \varepsilon_k) \ln(v_k) - \frac{1}{2} \sum_{k=1}^K \sum_{j=1}^K \beta_{kj} \ln(v_k) \ln(v_j). \quad (6)$$

The translog demand system represents a second order approximation to any functional form. Parametric restrictions imposed are adding up, along with a convenient normalization,

$$\sum_{k=1}^K \alpha_k = 1; \quad \sum_{k=1}^K \varepsilon_k = 0, \quad (7)$$

symmetry,

$$\beta_{kj} = \beta_{jk}, \quad \forall k, j, \quad (8)$$

and $\varepsilon \sim N_K(0, \Sigma)$. The latent budget share equations are

$$s_k^* \equiv v_k q_k = \frac{\alpha_k + \sum_{j=1}^K \beta_{kj} \ln(v_j) + \varepsilon_k}{D}, \quad k = 1, \dots, K \quad (9)$$

where

$$D \equiv 1 + \sum_{i=1}^K \sum_{j=1}^K \beta_{ij} \ln(v_j) > 0 \quad (10)$$

is strictly positive given a positive marginal utility of income. In contrast, the observed budget share equations are

$$s_k \equiv v_k x_k = \frac{\alpha_k + \sum_{j=1}^K \beta_{kj} \ln(\pi_j) + \varepsilon_k}{D^*}, \quad k = 1, \dots, K \quad (11)$$

Table 1
Summary statistics for three simulated data sets

	Number of observations		
	Data set 1	Data set 2	Data set 3
Panel A			
Good 1	641 12.82%	1192 23.84%	1548 30.96%
Good 2	369 7.38%	848 16.96%	1207 24.14%
Good 3	340 6.80%	902 18.04%	1336 26.72%
Good 4	199 3.98%	642 12.84%	996 19.92%
Good 5	108 2.16%	397 7.94%	677 13.54%
Good 6	660 13.20%	1155 23.10%	1489 29.78%
Panel B			
0 Corners	2919 58.38%	1308 26.16%	627 12.54%
1 Corner	1849 36.98%	2389 47.78%	2007 40.14%
2 Corners	228 4.56%	1163 23.26%	1873 37.46%
3 Corners	4 0.08%	139 2.78%	472 9.44%
4 Corners	0 0.00%	1 0.02%	21 0.42%
5 Corners	0 0.00%	0 0.00%	0 0.00%

Notes: Panel A reports the number of observations and sample proportion not consuming each good. Panel B breaks down the sample by the number of non-consumed goods. Total sample size is 5000 in each data set.

Table 2
Coherency results: three simulated data sets

Data set	Local coherency				Global coherency
	Mean	Median	SD	Max	
1	0.01	0	0.07	1	0.38
2	0.10	0	0.33	4	0.46
3	0.21	0	0.52	5	0.56

Notes: Mean and median reflect the number of draws of β rejected per iteration (after discarding the first 100 iterations). SD = standard deviation of rejections across the 9900 retained iterations. Max = maximum number of rejections during one iteration after discarding the first 100 iterations. Global coherency gives the percentage of β draws that are locally, but not globally, coherent.

where s_k equals zero if good k is not consumed (greater than zero and less than or equal to one otherwise), $\pi_j = \pi_j(\bar{v})$ is less than or equal to v_j if good j is not consumed (equal to v_j otherwise), and $D^* \equiv 1 + \sum_{i=1}^K \sum_{j=1}^K \beta_{ij} \ln(\pi_j)$.

For illustration, consider the regime where the first good is not consumed and the remaining $K - 1$ goods are consumed. Using (3) the log virtual price for first good is given by

$$\begin{aligned} \ln(\pi_1) &= -\frac{1}{\beta_{11}} \left[\alpha_1 + \sum_{j=2}^K \beta_{1j} \ln(v_j) + \varepsilon_1 \right] \\ &= \ln[\pi_1(v_2, \dots, v_K)]. \end{aligned} \quad (12)$$

While π_1 is unobserved, (5) implies that $\ln[\pi_1(v_2, \dots, v_K)] \leq \ln(v_1)$. In addition, since (12) can be re-written as

$$\ln(\pi_1) = \ln(v_1) - (1/\beta_{11})Ds_1^*, \quad (13)$$

this regime condition is equivalent to

$$\frac{1}{\beta_{11}}Ds_1^* \geq 0. \quad (14)$$

This regime is further characterized by restrictions on the latent budget shares of the consumed goods, s_k^* , $k = 2, \dots, K$. These

Table 3
Summary results for parameter estimates: data set #1

Parameter	Actual	Mean	SD	IQR	P1	P5	Median	P95	P99
$\alpha 1$	0.13	0.130	0.002	0.002	0.126	0.127	0.130	0.132	0.133
$\alpha 2$	0.15	0.149	0.002	0.002	0.146	0.147	0.150	0.152	0.153
$\alpha 3$	0.17	0.172	0.002	0.002	0.169	0.170	0.172	0.175	0.176
$\alpha 4$	0.19	0.189	0.003	0.002	0.185	0.186	0.189	0.192	0.193
$\alpha 5$	0.21	0.211	0.003	0.002	0.208	0.209	0.211	0.214	0.215
$\beta 11$	-0.40	-0.404	0.020	0.025	-0.446	-0.434	-0.403	-0.373	-0.362
$\beta 22$	-0.25	-0.261	0.021	0.026	-0.305	-0.292	-0.261	-0.229	-0.216
$\beta 33$	-0.50	-0.521	0.023	0.029	-0.572	-0.557	-0.521	-0.486	-0.471
$\beta 44$	-0.21	-0.237	0.023	0.032	-0.292	-0.276	-0.237	-0.199	-0.185
$\beta 55$	-0.18	-0.226	0.026	0.034	-0.285	-0.268	-0.226	-0.183	-0.167
$\beta 66$	-0.30	-0.300	0.024	0.032	-0.357	-0.340	-0.300	-0.262	-0.245
$\beta 12$	0.07	0.059	0.014	0.019	0.026	0.036	0.059	0.083	0.093
$\beta 13$	0.06	0.038	0.015	0.020	0.002	0.014	0.038	0.063	0.074
$\beta 14$	-0.02	-0.017	0.016	0.021	-0.053	-0.043	-0.016	0.010	0.020
$\beta 15$	0.08	0.053	0.017	0.023	0.014	0.025	0.053	0.080	0.092
$\beta 16$	0.10	0.104	0.016	0.021	0.068	0.078	0.104	0.129	0.139
$\beta 23$	0.09	0.080	0.016	0.022	0.043	0.054	0.080	0.107	0.117
$\beta 24$	0.05	0.041	0.016	0.022	0.002	0.014	0.041	0.068	0.079
$\beta 25$	-0.11	-0.124	0.018	0.024	-0.166	-0.153	-0.124	-0.094	-0.083
$\beta 26$	0.02	0.002	0.016	0.022	-0.035	-0.025	0.002	0.028	0.038
$\beta 34$	0.15	0.132	0.018	0.024	0.090	0.102	0.132	0.162	0.174
$\beta 35$	0.07	0.034	0.019	0.026	-0.012	0.002	0.034	0.066	0.078
$\beta 36$	0.03	0.025	0.017	0.023	-0.016	-0.003	0.025	0.052	0.064
$\beta 45$	-0.07	-0.091	0.020	0.027	-0.136	-0.123	-0.091	-0.057	-0.045
$\beta 46$	0.09	0.088	0.017	0.024	0.048	0.060	0.088	0.117	0.128
$\beta 56$	-0.05	-0.069	0.019	0.026	-0.113	-0.100	-0.069	-0.038	-0.027
$\sigma 11$	0.01	0.011	0.001	0.000	0.010	0.010	0.011	0.011	0.011
$\sigma 22$	0.01	0.010	0.001	0.000	0.009	0.010	0.010	0.010	0.010
$\sigma 33$	0.01	0.010	0.001	0.000	0.010	0.010	0.010	0.011	0.011
$\sigma 44$	0.01	0.011	0.001	0.000	0.010	0.010	0.011	0.011	0.011
$\sigma 55$	0.01	0.010	0.001	0.000	0.010	0.010	0.010	0.011	0.011
$\rho 12$	-0.30	-0.289	0.015	0.020	-0.323	-0.313	-0.289	-0.265	-0.255
$\rho 13$	-0.20	-0.197	0.016	0.021	-0.232	-0.222	-0.197	-0.172	-0.161
$\rho 14$	-0.10	-0.099	0.016	0.021	-0.136	-0.125	-0.100	-0.074	-0.062
$\rho 15$	-0.15	-0.157	0.015	0.021	-0.192	-0.182	-0.157	-0.132	-0.122
$\rho 23$	-0.10	-0.110	0.016	0.021	-0.147	-0.136	-0.110	-0.085	-0.075
$\rho 24$	-0.25	-0.237	0.015	0.020	-0.270	-0.261	-0.236	-0.212	-0.202
$\rho 25$	-0.17	-0.170	0.015	0.020	-0.204	-0.194	-0.170	-0.145	-0.134
$\rho 34$	-0.12	-0.114	0.016	0.020	-0.149	-0.139	-0.114	-0.088	-0.078
$\rho 35$	-0.10	-0.105	0.015	0.020	-0.140	-0.130	-0.105	-0.081	-0.071
$\rho 45$	-0.18	-0.206	0.015	0.019	-0.239	-0.229	-0.206	-0.181	-0.171

Notes: Results based on 10,000 iterations, discarding the initial 100. Local coherency checked. SD = standard deviation; IQR = interquartile range; P1, P5, P95, P99 = the 1st, 5th, 95th, and 99th percentiles, respectively.

3. Estimation

3.1. Issues

In the context of either the non-linear or linear translog demand system with many binding non-negativity constraints, there are two issues confronted during estimation. Firstly, if the model is estimated by maximum likelihood, the likelihood function for a particular observation requires that the random components for the non-consumed goods be integrated out (since the virtual prices are unobserved), implying as many integrals as non-consumed goods (e.g. Lee and Pitt (1986)). For large demand systems, with many non-consumed goods for some observations, this becomes taxing.

While alternative techniques have been developed to circumvent the dimensionality problem (e.g. Kao et al. (2001)), a second issue arises. Specifically, when estimating flexible functional forms of consumer or producer demands, imposing the theoretical curvature restrictions is crucial if the data contain zero demands. Even if the true data-generating process is coherent, one may still obtain inconsistent estimates of the parameters of the model if the iterative estimation process leaves the coherent region of the parameter space (van Soest et al., 1993). In the notation from above, statistical coherency requires: (i) that every possible vector ε of random components generates a unique set of observed demands,

x , and (ii) that every set of observed demands, x , can be generated by some ε vector. In practice, these conditions imply that the matrix of second-order partial derivatives of the cost or indirect utility function with respect to prices is negative semidefinite. This is also equivalent to the requirement that the matrix of Allen–Uzawa elasticities of substitution is negative semidefinite (van Soest and Kooreman, 1990).

A sufficient condition for the non-linear translog model to be locally coherent is that the indirect utility function be concave in the neighborhood of the observed budget shares. This requires that

$$ss' - \Delta(s) + \frac{1}{D - y'\beta e} [\beta - s(\beta e)' - \beta e s' + e'\beta e s'] \quad (23)$$

be a negative semidefinite matrix at each observed data point, where $\beta \equiv \{\beta_{ij}\}_{k,j=1}^K$ is a $K \times K$ matrix of slope parameters, $\Delta(s)$ is a $K \times K$ diagonal matrix with the observed shares s_k , $k = 1, \dots, K$, along the diagonal, s is a K -dimensional vector of observed shares, y is a K -dimensional vector of the differences between log market and virtual prices (i.e., with representative element $y_k = \ln(v_k) - \ln(\pi_k)$), and e is vector of ones defined previously (van Soest and Kooreman, 1990). In contrast, sufficient conditions for global coherency are that β is a negative definite matrix, $D > 0$, and $\beta e \leq 0$ (van Soest and Kooreman, 1990). While the conditions for global coherency are easier to impose since they are not data-specific, in practice they may be significantly more restrictive

Table 4

Summary results for parameter estimates: data set #2

Parameter	Actual	Mean	SD	IQR	P1	P5	Median	P95	P99
α_1	0.13	0.131	0.003	0.003	0.125	0.127	0.131	0.135	0.136
α_2	0.15	0.149	0.003	0.003	0.144	0.145	0.149	0.153	0.155
α_3	0.17	0.173	0.003	0.003	0.168	0.169	0.173	0.177	0.179
α_4	0.19	0.189	0.003	0.003	0.183	0.185	0.189	0.193	0.194
α_5	0.21	0.210	0.004	0.003	0.204	0.206	0.211	0.215	0.216
β_{11}	-0.40	-0.388	0.017	0.017	-0.419	-0.410	-0.388	-0.367	-0.358
β_{22}	-0.25	-0.248	0.014	0.017	-0.277	-0.268	-0.248	-0.227	-0.219
β_{33}	-0.50	-0.505	0.018	0.020	-0.539	-0.529	-0.505	-0.481	-0.472
β_{44}	-0.21	-0.220	0.015	0.019	-0.254	-0.243	-0.220	-0.197	-0.188
β_{55}	-0.18	-0.198	0.017	0.020	-0.236	-0.224	-0.198	-0.173	-0.163
β_{66}	-0.30	-0.300	0.017	0.021	-0.337	-0.326	-0.299	-0.274	-0.262
β_{12}	0.07	0.068	0.009	0.012	0.047	0.053	0.068	0.084	0.090
β_{13}	0.06	0.047	0.010	0.013	0.024	0.032	0.048	0.063	0.069
β_{14}	-0.02	-0.006	0.010	0.013	-0.030	-0.023	-0.006	0.010	0.016
β_{15}	0.08	0.066	0.011	0.014	0.042	0.050	0.067	0.083	0.090
β_{16}	0.10	0.105	0.011	0.014	0.080	0.087	0.105	0.122	0.130
β_{23}	0.09	0.086	0.010	0.013	0.063	0.070	0.086	0.102	0.109
β_{24}	0.05	0.051	0.010	0.013	0.028	0.035	0.051	0.068	0.074
β_{25}	-0.11	-0.110	0.011	0.013	-0.134	-0.127	-0.110	-0.093	-0.086
β_{26}	0.02	0.009	0.010	0.014	-0.015	-0.008	0.009	0.026	0.033
β_{34}	0.15	0.140	0.011	0.015	0.114	0.122	0.140	0.158	0.166
β_{35}	0.07	0.047	0.011	0.015	0.021	0.029	0.047	0.065	0.072
β_{36}	0.03	0.032	0.011	0.014	0.008	0.015	0.032	0.049	0.057
β_{45}	-0.07	-0.074	0.011	0.015	-0.099	-0.091	-0.074	-0.056	-0.049
β_{46}	0.09	0.095	0.011	0.015	0.069	0.077	0.095	0.112	0.120
β_{56}	-0.05	-0.056	0.012	0.015	-0.084	-0.075	-0.056	-0.037	-0.029
σ_{11}	0.02	0.021	0.003	0.001	0.020	0.020	0.021	0.022	0.022
σ_{22}	0.02	0.020	0.003	0.001	0.019	0.019	0.020	0.021	0.021
σ_{33}	0.02	0.021	0.002	0.001	0.019	0.020	0.021	0.022	0.022
σ_{44}	0.02	0.022	0.002	0.001	0.021	0.021	0.022	0.023	0.023
σ_{55}	0.02	0.021	0.002	0.001	0.019	0.020	0.021	0.022	0.022
ρ_{12}	-0.30	-0.284	0.017	0.022	-0.322	-0.311	-0.284	-0.257	-0.245
ρ_{13}	-0.20	-0.186	0.017	0.023	-0.225	-0.213	-0.186	-0.157	-0.144
ρ_{14}	-0.10	-0.094	0.018	0.024	-0.135	-0.123	-0.094	-0.065	-0.051
ρ_{15}	-0.15	-0.166	0.017	0.022	-0.205	-0.194	-0.166	-0.139	-0.127
ρ_{23}	-0.10	-0.117	0.017	0.023	-0.157	-0.146	-0.117	-0.089	-0.077
ρ_{24}	-0.25	-0.241	0.017	0.023	-0.279	-0.269	-0.241	-0.213	-0.201
ρ_{25}	-0.17	-0.162	0.018	0.023	-0.202	-0.191	-0.162	-0.134	-0.121
ρ_{34}	-0.12	-0.123	0.018	0.025	-0.166	-0.153	-0.123	-0.093	-0.081
ρ_{35}	-0.10	-0.122	0.016	0.022	-0.160	-0.149	-0.123	-0.095	-0.084
ρ_{45}	-0.18	-0.202	0.018	0.024	-0.242	-0.230	-0.202	-0.173	-0.160

Notes: Results based on 10,000 iterations, discarding the initial 100. Local coherency checked. SD = standard deviation; IQR = interquartile range; P1, P5, P95, P99 = the 1st, 5th, 95th, and 99th percentiles, respectively.

than necessary. Of course, within our estimation approach, this restrictiveness can be 'measured' in practice by comparing results under the imposition of global versus local coherency.

In the linear translog model, $\beta e = 0, D$ reduces to one, and the local coherency condition in (23) simplifies to the requirement that $s' - \Delta(s) + \beta$

be a negative semidefinite matrix for each observation (Diewert and Wales, 1987). Global coherency only requires that β be a negative definite matrix.¹

3.2. Bayesian estimation

To circumvent the dimensionality problem, as well as impose only the minimum necessary requirement for coherency, we follow Pitt and Millimet (1999) and utilize the Markov Chain Monte Carlo (MCMC) technique of Gibbs' Sampling combined with data augmentation. In models with a straightforward latent structure, the observed data may be augmented in order to

facilitate application of the Gibbs sampler (Tanner and Wong 1987). In the present application, this process converts the observed demands into latent demands subject to the appropriate regime conditions. The ability to augment the data to include the latent budget shares eliminates the need for high-dimensional integration since the latent demands depend only on the observed market prices. Furthermore, as noted in Terrell (1996), O'Donnell et al. (1999), and Pitt and Millimet (1999), rejection sampling offers a straightforward mechanism to impose the conditions needed for local coherency. Specifically, during each iteration of the Gibbs sampler, it is straightforward to reject parameter draws that violate the concavity condition for even a single observation. Thus, the Gibbs Sampling technique solves both the dimensionality problem and the problem of how to impose only local coherency.

The desired density to be sampled from is $p(\alpha, \beta, \Sigma|s, v)$, corresponding to the random (stochastic) indirect utility function in (6), where s is the vector of observed budget shares. After augmenting the data, the actual density sampled from is $p(\alpha, \beta, \Sigma|s^*, v)$, where s^* is a vector of latent budget shares.²

¹ Terrell (1996) and O'Donnell et al. (1999) impose regularity conditions for a range of prices, rather than globally or only at the observed data points, by evaluating these conditions at all prices in a small grid around actual prices. This represents a trivial extension to ensuring the model is locally coherent.

² To be clear, once the data has been augmented to include s^* , we sample from $p(\alpha, \beta, \Sigma|s^*)$, thereby omitting s from the conditioning set. This is because, conditional on the augmented data, the original data contain no additional information.

Table 5
Summary results for parameter estimates: data set #3

Parameter	Actual	Mean	SD	IQR	P1	P5	Median	P95	P99
$\alpha 1$	0.13	0.133	0.004	0.004	0.125	0.128	0.133	0.138	0.140
$\alpha 2$	0.15	0.149	0.003	0.004	0.142	0.144	0.149	0.154	0.156
$\alpha 3$	0.17	0.174	0.004	0.004	0.167	0.169	0.174	0.180	0.182
$\alpha 4$	0.19	0.189	0.004	0.004	0.181	0.183	0.189	0.194	0.196
$\alpha 5$	0.21	0.209	0.004	0.005	0.201	0.203	0.209	0.214	0.216
$\beta 11$	-0.40	-0.379	0.015	0.015	-0.406	-0.398	-0.379	-0.360	-0.354
$\beta 22$	-0.25	-0.245	0.011	0.014	-0.268	-0.261	-0.245	-0.228	-0.221
$\beta 33$	-0.50	-0.497	0.017	0.017	-0.527	-0.517	-0.497	-0.477	-0.469
$\beta 44$	-0.21	-0.214	0.012	0.015	-0.240	-0.233	-0.214	-0.196	-0.189
$\beta 55$	-0.18	-0.192	0.013	0.016	-0.220	-0.211	-0.192	-0.173	-0.166
$\beta 66$	-0.30	-0.302	0.014	0.017	-0.332	-0.323	-0.301	-0.280	-0.272
$\beta 12$	0.07	0.069	0.008	0.010	0.051	0.057	0.069	0.082	0.087
$\beta 13$	0.06	0.050	0.008	0.011	0.031	0.037	0.050	0.064	0.069
$\beta 14$	-0.02	0.000	0.008	0.011	-0.020	-0.014	0.000	0.013	0.019
$\beta 15$	0.08	0.074	0.008	0.011	0.055	0.061	0.074	0.087	0.093
$\beta 16$	0.10	0.106	0.009	0.012	0.085	0.091	0.106	0.120	0.126
$\beta 23$	0.09	0.085	0.008	0.011	0.066	0.072	0.085	0.098	0.104
$\beta 24$	0.05	0.052	0.008	0.010	0.034	0.040	0.053	0.065	0.070
$\beta 25$	-0.11	-0.106	0.008	0.010	-0.124	-0.119	-0.106	-0.094	-0.088
$\beta 26$	0.02	0.014	0.008	0.011	-0.006	0.001	0.014	0.028	0.033
$\beta 34$	0.15	0.145	0.009	0.012	0.124	0.131	0.145	0.159	0.165
$\beta 35$	0.07	0.055	0.009	0.011	0.035	0.041	0.055	0.069	0.075
$\beta 36$	0.03	0.033	0.009	0.012	0.013	0.019	0.033	0.048	0.055
$\beta 45$	-0.07	-0.070	0.009	0.011	-0.090	-0.084	-0.070	-0.057	-0.051
$\beta 46$	0.09	0.096	0.009	0.012	0.076	0.082	0.096	0.110	0.117
$\beta 56$	-0.05	-0.053	0.009	0.012	-0.074	-0.068	-0.053	-0.038	-0.032
$\sigma 11$	0.03	0.032	0.004	0.001	0.030	0.031	0.032	0.034	0.034
$\sigma 22$	0.03	0.030	0.004	0.001	0.028	0.028	0.030	0.031	0.032
$\sigma 33$	0.03	0.032	0.004	0.001	0.030	0.030	0.032	0.034	0.034
$\sigma 44$	0.03	0.033	0.003	0.001	0.031	0.032	0.033	0.035	0.036
$\sigma 55$	0.03	0.031	0.004	0.001	0.029	0.029	0.031	0.032	0.033
$\rho 12$	-0.30	-0.287	0.018	0.024	-0.328	-0.316	-0.287	-0.258	-0.245
$\rho 14$	-0.20	-0.175	0.018	0.025	-0.218	-0.206	-0.175	-0.145	-0.133
$\rho 14$	-0.10	-0.093	0.020	0.027	-0.139	-0.125	-0.093	-0.061	-0.047
$\rho 15$	-0.15	-0.183	0.018	0.024	-0.226	-0.212	-0.183	-0.154	-0.142
$\rho 23$	-0.10	-0.119	0.019	0.025	-0.161	-0.149	-0.118	-0.088	-0.076
$\rho 24$	-0.25	-0.238	0.018	0.024	-0.280	-0.268	-0.238	-0.208	-0.196
$\rho 25$	-0.17	-0.148	0.019	0.025	-0.190	-0.178	-0.149	-0.117	-0.104
$\rho 34$	-0.12	-0.144	0.020	0.027	-0.191	-0.177	-0.144	-0.111	-0.099
$\rho 35$	-0.10	-0.142	0.018	0.024	-0.183	-0.171	-0.142	-0.114	-0.102
$\rho 45$	-0.18	-0.181	0.020	0.027	-0.226	-0.213	-0.181	-0.148	-0.134

Notes: Results based on 10,000 iterations, discarding the initial 100. Local coherency checked. SD = standard deviation; IQR = interquartile range; P1, P5, P95, P99 = the 1st, 5th, 95th, and 99th percentiles, respectively.

This density may be sampled by cycling through the following conditional densities:

- (i) $p(s^*|\alpha, \beta, \Sigma, s, v)$
- (ii) $p(\beta|\alpha, s^*, v, \Sigma)$
- (iii) $p(\alpha|\beta, s^*, v, \Sigma)$
- (iv) $p(\Sigma|s^*, v, \alpha, \beta)$.

The Gibbs sampler constructs sequences of draws by sampling consecutively from the above conditional distributions. We now discuss the precise steps.

3.2.1. Augmenting budget shares

The first step is to augment the data by sampling the latent budget shares conditional on the observed data and initial values for the parameters of the model (α , β , and Σ). The algorithm proposed here represents a significant departure from the algorithm in Pitt and Millimet (1999). Pitt and Millimet (1999) begin by sampling the random components, ε , for the non-consumed goods subject to truncation points defined implicitly by (17). After sampling these error components, the authors proceed to sample the error components for the consumed goods conditional on the error components for the non-consumed goods subject to the regime conditions given in (19).

Here, we simplify matters in two ways. Firstly, rather than sampling ε , we sample Z from the non-negative portion of the

normal distribution as shown in (17). After sampling Z , we are able to back out the latent budget shares of the non-consumed goods, s_k^* , as well as the corresponding virtual prices, $\ln(\pi_k)$. Formally, for the general case where the first K_0 goods are not consumed for a particular observation,

$$Z = B^{-1} \begin{bmatrix} Ds_1^* \\ \vdots \\ Ds_{K_0}^* \end{bmatrix} \sim N_{K_0}^+(0, B^{-1}D^2\Sigma B^{-1}).$$

Given Z , $s_1^*, \dots, s_{K_0}^*$ are obtained using

$$\begin{bmatrix} s_1^* \\ \vdots \\ s_{K_0}^* \end{bmatrix} = \left(\frac{1}{D}\right) BZ \quad (24)$$

and $\ln(\pi_1), \dots, \ln(\pi_{K_0})$ are obtained using (16).

Our second simplification relative to Pitt and Millimet (1999) is that we do not sample the latent budget shares for the consumed goods conditional on the preceding latent budget shares for the non-consumed goods. Rather, we make use of the relationship we noted above in (20) and solve for $s_{K_0+1}^*, \dots, s_K^*$ deterministically given $s_1^*, \dots, s_{K_0}^*$, initial values of the parameters, and the observed data. This ensures the regime conditions in (19) will be met, and

Table 6
Summary statistics for three simulated data sets

	Number of observations		
	Data set 1	Data set 2	Data set 3
Panel A			
Good 1	654 13.08%	1260 25.20%	1681 33.62%
Good 2	399 7.98%	957 19.14%	1335 26.70%
Good 3	341 6.82%	953 19.06%	1442 28.84%
Good 4	198 3.96%	630 12.60%	1001 20.02%
Good 5	124 2.48%	483 9.66%	854 17.08%
Good 6	881 17.62%	1582 31.64%	1990 39.80%
Panel B			
0 Corners	2718 54.36%	1063 21.26%	441 8.82%
1 Corner	1974 39.48%	2279 45.58%	1771 35.42%
2 Corners	301 6.02%	1395 27.90%	1918 38.36%
3 Corners	7 0.14%	256 5.12%	788 15.76%
4 Corners	0 0.00%	7 0.14%	78 1.56%
5 Corners	0 0.00%	0 0.00%	4 0.08%

Notes: Panel A reports the number of observations and sample proportion not consuming each good. Panel B breaks down the sample by the number of non-consumed goods. Total sample size is 5000 in each data set.

thus reduces the computational burden significantly relative to Pitt and Millimet (1999).

For clarity, for the general case where the first K_0 goods are not consumed for a particular observation, the m th iteration of the Gibbs sampler algorithm entails:

- (ABS-i) draw $Z_1^{(m)}, \dots, Z_{K_0}^{(m)} \sim N_{K_0}^+(0, B_{(m-1)}^{-1} D_{(m-1)}^2 \Sigma_{(m-1)} B_{(m-1)}^{-1})$
- (ABS-ii) obtain $s_1^{*(m)}, \dots, s_{K_0}^{*(m)}$ using (24)
- (ABS-iii) obtain $\ln(\pi_1^{(m)}), \dots, \ln(\pi_{K_0}^{(m)})$ using (16)
- (ABS-iv) obtain $s_{K_0+1}^{*(m)}, \dots, s_K^{*(m)}$ using (20).

3.2.2. Remainder of the algorithm

3.2.2.1. Non-linear translog. The restrictions in (7) result in Σ being singular. Thus, only the first $K - 1$ equations are used in the estimation of α , β , and Σ , with the results being invariant to which good is omitted. While this leads one to sample only the leading $K - 1$ elements of α and $(K - 1) \times (K - 1)$ elements of Σ , we can sample all of the elements of β as we now describe.

After augmenting the data with the latent budget shares, we draw updated values of α and β from the appropriate distributions subject to the local coherency condition in (23) using rejection sampling. Here, again, we diverge from Pitt and Millimet (1999), who utilize non-linear SUR to derive the appropriate distributions from which to sample α and β . Rather than follow suit, which requires an iterative estimation algorithm within each iteration of the Gibbs Sampler, we utilize an auxiliary regression as is described below.

To proceed, we transform the latent share equations into a linear form by substituting (10) into (9) and re-arranging to obtain

$$s_k^* = \alpha_k + \sum_j \left[\beta_{kj} - s_k^* \sum_i \beta_{ij} \right] \ln(v_j) + \varepsilon_k, \quad k = 1, \dots, K - 1.$$

Table 7
Coherency results: three simulated data sets

Data Set	Local coherency				Global coherency
	Mean	Median	SD	Max	
1	0.00	0	0.00	0	0.0004
2	0.00	0	0.00	0	0.00
3	0.00	0	0.00	0	0.00

Notes: Mean and median reflect the number of draws of β rejected per iteration (after discarding the first 100 iterations). SD = standard deviation of rejections across the 9900 retained iterations. Max = maximum number of rejections during one iteration after discarding the first 100 iterations. Global coherency gives the percentage of β draws that are locally, but not globally, coherent.

However, the inclusion of $s_k^* \ln(v_j)$, $j = 1, \dots, K$, as covariates creates a problem since these terms are correlated with the error term. To circumvent this problem, we replace the covariates with their linear projection on a vector of squared log market prices and a full set of interaction terms between log market prices (Lewbel, 1997), which we denote as R .

We can rewrite the model in (9) in matrix notation as follows

$$S^* = (\alpha + V\beta_v + \varepsilon)D^{-1}, \tag{25}$$

where S^* , α , and ε are now $(K - 1) \times 1$ vectors of s_k^* , α_k , and ε_k , respectively; β_v is a $(K(K + 1)/2) \times 1$ column vector containing the unique elements of β , and V is a matrix of log normalized market prices arranged appropriately. Note, in this notation D is equal to $1 + e'V\beta_v$. Rearranging terms leads to the following:

$$Y = X\beta_v + \varepsilon,$$

where $Y = S^* - \alpha$ and $X = (V - S^*e'V)$. If the prior distribution of β_v is $N(b_0, B_0)$, and denoting $Z = R(R'R)^{-1}R'X$, then the conditional posterior distribution of β_v is $p(\beta_v | \alpha, s^*, v, \Sigma) \sim N(B_\beta b_\beta, B_\beta)$, where

$$B_\beta = \left(\sum_n Z' \Sigma^{-1} Z + B_0^{-1} \right)^{-1}$$

$$b_\beta = \sum_n Z' \Sigma^{-1} Y + B_0^{-1} b_0,$$

$n = 1, \dots, N$ indexes observations, and Σ now represents the $(K - 1) \times (K - 1)$ variance-covariance matrix of ε .

Rearranging terms in (25) and assuming the prior distribution of α to be $N(a_0, A_0)$, we obtain the conditional posterior distribution of α as $p(\alpha | \beta, s^*, v, \Sigma) \sim N(B_\alpha b_\alpha, B_\alpha)$, where

$$B_\alpha = \left(\sum_n X'_\alpha \Sigma^{-1} X_\alpha + A_0^{-1} \right)^{-1}$$

$$b_\alpha = \sum_n X'_\alpha \Sigma^{-1} Y_\alpha + A_0^{-1} a_0,$$

where $Y_\alpha = S^*D - V\beta_v$ and X_α is an identity matrix.

After sampling locally coherent values of α and β , the final step entails updating Σ . An updated value is obtained by drawing Σ^{-1} from an Wishart density. Assuming the prior distribution of Σ^{-1} is $IW(S_0, s_0)$, the conditional posterior is given by $p(\Sigma | s^*, v, \alpha, \beta) \sim IW((S_0^{-1} + \sum_n E'E)^{-1}, N + s_0)$, where $E = S^* - \alpha - V\beta_v$.

For clarity, the m th iteration of the Gibbs sampler algorithm entails:

- (NL-i) Sample the $K(K + 1)/2$ unique elements of β . Check coherency (either local or global for comparison).
- (NL-ii) Sample the $(K - 1) \times 1$ sub-vector of α and compute $\alpha_1, \dots, \alpha_{K-1}$ and compute the remaining element.
- (NL-iii) Sample the $(K - 1) \times (K - 1)$ sub-matrix of Σ and compute the remaining elements.

Table 8
Summary results for parameter estimates: data set #1

Parameter	Actual	Mean	SD	IQR	P1	P5	Median	P95	P99
α_1	0.13	0.130	0.002	0.002	0.126	0.127	0.130	0.132	0.133
α_2	0.15	0.149	0.002	0.002	0.146	0.147	0.149	0.152	0.153
α_3	0.17	0.173	0.002	0.002	0.169	0.170	0.173	0.175	0.176
α_4	0.19	0.189	0.002	0.002	0.185	0.186	0.189	0.191	0.192
α_5	0.21	0.212	0.003	0.002	0.208	0.209	0.212	0.214	0.215
β_{11}	-0.40	-0.396	0.015	0.018	-0.428	-0.418	-0.396	-0.374	-0.365
β_{22}	-0.25	-0.251	0.013	0.017	-0.281	-0.272	-0.251	-0.229	-0.220
β_{33}	-0.50	-0.498	0.016	0.018	-0.529	-0.520	-0.498	-0.477	-0.468
β_{44}	-0.21	-0.226	0.014	0.018	-0.257	-0.248	-0.226	-0.203	-0.194
β_{55}	-0.18	-0.185	0.013	0.018	-0.216	-0.207	-0.185	-0.164	-0.154
β_{66}	-0.80	-0.803	0.024	0.024	-0.844	-0.832	-0.803	-0.773	-0.761
β_{12}	0.07	0.067	0.010	0.013	0.044	0.051	0.067	0.083	0.089
β_{13}	0.06	0.050	0.010	0.013	0.028	0.035	0.050	0.066	0.072
β_{14}	-0.02	-0.010	0.010	0.013	-0.032	-0.026	-0.010	0.006	0.012
β_{15}	0.08	0.074	0.010	0.013	0.052	0.058	0.074	0.090	0.097
β_{16}	0.21	0.215	0.012	0.015	0.189	0.197	0.215	0.233	0.241
β_{23}	0.09	0.097	0.009	0.012	0.075	0.081	0.096	0.112	0.118
β_{24}	0.05	0.052	0.010	0.013	0.030	0.036	0.052	0.068	0.074
β_{25}	-0.11	-0.102	0.010	0.013	-0.124	-0.117	-0.102	-0.087	-0.081
β_{26}	0.15	0.138	0.011	0.014	0.112	0.120	0.138	0.155	0.162
β_{34}	0.15	0.149	0.010	0.013	0.126	0.133	0.149	0.165	0.171
β_{35}	0.07	0.065	0.009	0.013	0.043	0.049	0.064	0.080	0.087
β_{36}	0.13	0.139	0.011	0.015	0.114	0.121	0.138	0.157	0.164
β_{45}	-0.07	-0.064	0.010	0.013	-0.086	-0.080	-0.064	-0.048	-0.042
β_{46}	0.10	0.099	0.011	0.015	0.073	0.081	0.099	0.117	0.125
β_{56}	0.21	0.212	0.011	0.015	0.187	0.194	0.212	0.231	0.238
σ_{11}	0.01	0.010	0.002	0.000	0.010	0.010	0.010	0.011	0.011
σ_{22}	0.01	0.010	0.001	0.000	0.009	0.010	0.010	0.010	0.010
σ_{33}	0.01	0.010	0.001	0.000	0.010	0.010	0.010	0.010	0.011
σ_{44}	0.01	0.010	0.001	0.000	0.010	0.010	0.010	0.011	0.011
σ_{55}	0.01	0.010	0.001	0.000	0.010	0.010	0.010	0.011	0.011
ρ_{12}	-0.30	-0.285	0.015	0.020	-0.318	-0.309	-0.285	-0.261	-0.251
ρ_{13}	-0.20	-0.199	0.015	0.020	-0.233	-0.223	-0.199	-0.174	-0.164
ρ_{14}	-0.10	-0.104	0.016	0.021	-0.140	-0.129	-0.104	-0.079	-0.068
ρ_{15}	-0.15	-0.152	0.015	0.020	-0.187	-0.177	-0.152	-0.127	-0.117
ρ_{23}	-0.10	-0.108	0.015	0.021	-0.145	-0.134	-0.108	-0.083	-0.072
ρ_{24}	-0.25	-0.235	0.015	0.020	-0.269	-0.259	-0.235	-0.211	-0.201
ρ_{25}	-0.17	-0.157	0.016	0.021	-0.193	-0.183	-0.157	-0.132	-0.120
ρ_{34}	-0.12	-0.123	0.015	0.020	-0.158	-0.148	-0.123	-0.099	-0.089
ρ_{35}	-0.10	-0.100	0.015	0.020	-0.134	-0.124	-0.100	-0.076	-0.066
ρ_{45}	-0.18	-0.200	0.015	0.020	-0.234	-0.224	-0.200	-0.176	-0.166

Notes: Results based on 10,000 iterations, discarding the initial 100. Local coherency checked. SD = standard deviation; IQR = interquartile range; P1, P5, P95, P99 = the 1st, 5th, 95th, and 99th percentiles, respectively.

3.2.2.2. *Linear translog*. Similarly, the linear translog demand system uses a seemingly unrelated regression (SUR) approach, imposing the cross-equation restrictions implied by adding up and symmetry; see (7) and (8). To impose $\beta_{ke} = 0 \forall k$, we only estimate the off-diagonal elements of β and set diagonal elements to the negative sum of the off-diagonal elements in each row. This can be achieved by using differences in log prices in the estimation of each equation since

$$s_k^* = \alpha_k + \ln(v_k) \left(- \sum_{j \neq k} \beta_{jk} \right) + \sum_{j \neq k} \ln(v_j) \beta_{kj} + \varepsilon_k$$

can be written as

$$s_k^* = \alpha_k + \sum_{j \neq k} [\ln(v_j) - \ln(v_k)] \beta_{kj} + \varepsilon_k.$$

Hence, we sample only the $K(K-1)/2$ free off-diagonal parameters of β . The rest of the algorithm follows the same logic as in the case of the non-linear translog, except we no longer need to use auxiliary regression.

For clarity, the m th iteration of the Gibbs sampler algorithm entails:

(L-i) Sample the $K(K-1)/2$ unique elements of β . Check coherency (either local or global for comparison).

(L-ii) Sample the $(K-1) \times 1$ sub-vector of α and compute $\alpha_1, \dots, \alpha_{K-1}$ and compute the remaining element.

(L-iii) Sample the $(K-1) \times (K-1)$ sub-matrix of Σ and compute the remaining elements.

4. Simulations

To assess the performance of our proposed estimator, we simulate several data sets from the non-linear and linear translog models. Specifically, for each model, we generate three data sets, each containing 5000 observations consuming possibly six goods. Market prices for each good are drawn from independent normal distributions, with mean zero and standard deviations of 0.1, 0.2, or 0.3 across the three data sets. Errors are drawn from a multivariate normal distribution with mean zero and variances that vary across the three data sets. The true values of β are chosen such that the data-generating process is 'barely' globally coherent; the largest eigenvalue is -0.0034 , the largest element of βe is -0.01 , and D is positive for all observations. Adding up and symmetry hold in all simulated data sets.

Summary statistics for the three simulated non-linear translog models are given in Table 1. Data Set #1 refers to the data simulated with log normalized market prices that have a standard deviation of 0.1; Data Set #2 (#3) refer to the data simulated with log normalized market prices that have a standard deviation of 0.2 (0.3).

Table 9

Summary results for parameter estimates: data set #2

Parameter	Actual	Mean	SD	IQR	P1	P5	Median	P95	P99
α_1	0.13	0.131	0.003	0.003	0.126	0.127	0.131	0.135	0.136
α_2	0.15	0.148	0.003	0.003	0.143	0.144	0.148	0.151	0.153
α_3	0.17	0.175	0.003	0.003	0.170	0.171	0.175	0.179	0.180
α_4	0.19	0.189	0.003	0.003	0.184	0.185	0.189	0.192	0.194
α_5	0.21	0.212	0.003	0.003	0.207	0.208	0.212	0.216	0.217
β_{11}	-0.40	-0.393	0.014	0.014	-0.416	-0.409	-0.393	-0.376	-0.368
β_{22}	-0.25	-0.251	0.012	0.013	-0.274	-0.267	-0.251	-0.235	-0.228
β_{33}	-0.50	-0.495	0.015	0.014	-0.520	-0.511	-0.495	-0.478	-0.471
β_{44}	-0.21	-0.221	0.010	0.013	-0.243	-0.236	-0.221	-0.205	-0.198
β_{55}	-0.18	-0.187	0.010	0.013	-0.210	-0.203	-0.187	-0.171	-0.164
β_{66}	-0.80	-0.813	0.024	0.021	-0.849	-0.838	-0.812	-0.787	-0.777
β_{12}	0.07	0.069	0.008	0.010	0.052	0.057	0.069	0.081	0.086
β_{13}	0.06	0.047	0.008	0.010	0.030	0.035	0.047	0.059	0.065
β_{14}	-0.02	-0.010	0.007	0.010	-0.026	-0.022	-0.010	0.002	0.007
β_{15}	0.08	0.077	0.007	0.010	0.060	0.065	0.077	0.089	0.093
β_{16}	0.21	0.209	0.010	0.012	0.188	0.194	0.209	0.224	0.230
β_{23}	0.09	0.096	0.007	0.010	0.079	0.084	0.096	0.108	0.113
β_{24}	0.05	0.050	0.007	0.010	0.033	0.038	0.050	0.061	0.066
β_{25}	-0.11	-0.107	0.008	0.010	-0.124	-0.119	-0.107	-0.095	-0.091
β_{26}	0.15	0.144	0.009	0.012	0.124	0.129	0.144	0.158	0.164
β_{34}	0.15	0.146	0.008	0.010	0.130	0.135	0.146	0.158	0.163
β_{35}	0.07	0.064	0.007	0.010	0.047	0.052	0.064	0.076	0.081
β_{36}	0.13	0.141	0.010	0.012	0.120	0.126	0.141	0.156	0.162
β_{45}	-0.07	-0.066	0.007	0.010	-0.082	-0.077	-0.066	-0.054	-0.048
β_{46}	0.10	0.100	0.009	0.012	0.080	0.086	0.100	0.114	0.120
β_{56}	0.21	0.219	0.010	0.012	0.198	0.204	0.219	0.233	0.239
σ_{11}	0.02	0.021	0.003	0.001	0.020	0.020	0.021	0.022	0.022
σ_{22}	0.02	0.021	0.003	0.001	0.020	0.020	0.021	0.022	0.022
σ_{33}	0.02	0.021	0.003	0.001	0.019	0.020	0.021	0.021	0.022
σ_{44}	0.02	0.021	0.003	0.001	0.020	0.020	0.021	0.022	0.022
σ_{55}	0.02	0.021	0.003	0.001	0.020	0.021	0.021	0.022	0.023
ρ_{12}	-0.30	-0.286	0.017	0.023	-0.325	-0.313	-0.286	-0.258	-0.246
ρ_{13}	-0.20	-0.186	0.017	0.022	-0.226	-0.214	-0.186	-0.158	-0.146
ρ_{14}	-0.10	-0.100	0.018	0.024	-0.141	-0.129	-0.100	-0.071	-0.058
ρ_{15}	-0.15	-0.158	0.017	0.023	-0.198	-0.187	-0.159	-0.130	-0.118
ρ_{23}	-0.10	-0.128	0.017	0.023	-0.167	-0.156	-0.128	-0.099	-0.087
ρ_{24}	-0.25	-0.229	0.017	0.023	-0.268	-0.256	-0.229	-0.201	-0.188
ρ_{25}	-0.17	-0.119	0.021	0.026	-0.163	-0.151	-0.119	-0.087	-0.072
ρ_{34}	-0.12	-0.133	0.017	0.023	-0.172	-0.161	-0.133	-0.105	-0.093
ρ_{35}	-0.10	-0.119	0.017	0.023	-0.159	-0.147	-0.119	-0.091	-0.080
ρ_{45}	-0.18	-0.188	0.019	0.024	-0.229	-0.217	-0.188	-0.159	-0.145

Notes: Results based on 10,000 iterations, discarding the initial 100. Local coherency checked. SD = standard deviation; IQR = interquartile range; P1, P5, P95, P99 = the 1st, 5th, 95th, and 99th percentiles, respectively.

As can be seen, each data set contains a large fraction of non-consumed goods. In Data Set #1, over 40% of the sample has at least one non-consumed good, although only four observations consume less than four of the six goods. Moreover, the corners are fairly dispersed across the six goods, as each good is not consumed by at least 2% of the sample. In Data Sets #2 and #3, the number of binding non-negativity constraints is even higher, with over 73% and 87% of observations not consuming at least one good, respectively. Moreover, a sizeable fraction of households in each data set consume only three of the six goods, and a few consume only two goods.

For each of the three data sets, we iterate the Gibbs Sampler 10,000 times and discard the initial 100 iterations. Before looking at the parameter estimates for the non-linear translog models, the number of draws of β per iteration that fail to satisfy local coherency for each of the three data sets are given in Table 2. In addition, we also report the fraction of draws of β that satisfy local, but not global, coherency. The results indicate that, on average, roughly one draw of β every other iteration fails to satisfy local coherency. However, of the draws that satisfy local coherency, between 38% (Data Set #1) and 56% (Data Set #3) do not satisfy global coherency, despite the underlying data-generating process being globally coherent.

In terms of the actual parameter results, Tables 3–5 displays the results for the three data sets. The first column of numbers

in each table reports the true value used to simulate the data; the remaining values give various summary statistics of the distribution of locally coherent parameter draws after discarding the initial 100 iterations. In the interest of brevity, the results can be summarized easily: the results are encouraging. In all cases, the means of the posterior density of α and β lie within two standard deviations of the true value. While a few of the true values of the covariance parameters lie outside the corresponding interval, the estimates are still very close to the truth. While these results are only for three data sets, given that the proportion of corners in the simulated data, especially Data Set #3, is very high relative to most actual data sets, the results are promising.

Tables 6–10 are analogous to Tables 1–5 except the data sets are generating according to the linear translog model. Again, the sample fraction not consuming at least one good is quite high (see Table 6). Specifically, over 50% in Data Set #1, almost 80% in Data Set #2, and over 90% in Data Set #3 have at least one corner; over 15% have three or more corners in Data Set #3. That said, Table 7 indicates that all parameter draws satisfy both local or global coherency. Finally, the results in Tables 8–10 continue to look very good, particularly with respect to the α and β . Thus, our simulation results suggest that our methodology represents a potentially feasible way to recover consistent estimates of high-dimensional demand systems containing many binding non-negativity constraints.

Table 10
Summary results for parameter estimates: data set #3

Parameter	Actual	Mean	SD	IQR	P1	P5	Median	P95	P99
$\alpha 1$	0.13	0.133	0.003	0.004	0.126	0.128	0.133	0.138	0.140
$\alpha 2$	0.15	0.146	0.003	0.004	0.139	0.141	0.146	0.150	0.152
$\alpha 3$	0.17	0.176	0.003	0.004	0.169	0.172	0.176	0.181	0.183
$\alpha 4$	0.19	0.188	0.004	0.004	0.181	0.183	0.188	0.193	0.194
$\alpha 5$	0.21	0.210	0.004	0.004	0.203	0.205	0.210	0.215	0.217
$\beta 11$	-0.40	-0.395	0.014	0.013	-0.417	-0.411	-0.395	-0.380	-0.374
$\beta 22$	-0.25	-0.251	0.011	0.011	-0.271	-0.265	-0.251	-0.237	-0.231
$\beta 33$	-0.50	-0.491	0.015	0.013	-0.513	-0.506	-0.491	-0.475	-0.468
$\beta 44$	-0.21	-0.221	0.010	0.011	-0.241	-0.235	-0.221	-0.207	-0.202
$\beta 55$	-0.18	-0.191	0.009	0.011	-0.210	-0.205	-0.191	-0.176	-0.171
$\beta 66$	-0.80	-0.828	0.025	0.021	-0.865	-0.854	-0.828	-0.802	-0.792
$\beta 12$	0.07	0.068	0.007	0.009	0.053	0.057	0.068	0.079	0.084
$\beta 13$	0.06	0.049	0.007	0.009	0.034	0.038	0.049	0.060	0.065
$\beta 14$	-0.02	-0.010	0.006	0.009	-0.024	-0.020	-0.009	0.001	0.005
$\beta 15$	0.08	0.074	0.007	0.009	0.060	0.064	0.074	0.085	0.090
$\beta 16$	0.21	0.213	0.010	0.012	0.194	0.199	0.213	0.227	0.234
$\beta 23$	0.09	0.098	0.007	0.009	0.083	0.087	0.098	0.109	0.113
$\beta 24$	0.05	0.045	0.006	0.008	0.031	0.035	0.045	0.056	0.060
$\beta 25$	-0.11	-0.110	0.007	0.008	-0.124	-0.120	-0.109	-0.099	-0.095
$\beta 26$	0.15	0.149	0.009	0.011	0.131	0.136	0.149	0.162	0.168
$\beta 34$	0.15	0.148	0.007	0.009	0.133	0.137	0.148	0.158	0.163
$\beta 35$	0.07	0.068	0.007	0.009	0.053	0.058	0.068	0.079	0.084
$\beta 36$	0.13	0.127	0.009	0.012	0.108	0.113	0.127	0.142	0.148
$\beta 45$	-0.07	-0.072	0.006	0.008	-0.086	-0.082	-0.072	-0.062	-0.058
$\beta 46$	0.10	0.110	0.008	0.011	0.092	0.096	0.109	0.123	0.129
$\beta 56$	0.21	0.230	0.010	0.011	0.210	0.216	0.229	0.243	0.249
$\sigma 11$	0.03	0.032	0.005	0.001	0.030	0.031	0.032	0.034	0.035
$\sigma 22$	0.03	0.032	0.005	0.001	0.030	0.030	0.032	0.033	0.034
$\sigma 33$	0.03	0.032	0.005	0.001	0.030	0.031	0.032	0.034	0.035
$\sigma 44$	0.03	0.033	0.004	0.001	0.031	0.032	0.033	0.034	0.035
$\sigma 55$	0.03	0.034	0.004	0.001	0.032	0.033	0.034	0.036	0.036
$\rho 12$	-0.30	-0.286	0.018	0.024	-0.328	-0.315	-0.286	-0.256	-0.243
$\rho 13$	-0.20	-0.176	0.019	0.025	-0.219	-0.206	-0.176	-0.145	-0.133
$\rho 14$	-0.10	-0.101	0.019	0.026	-0.144	-0.132	-0.101	-0.069	-0.057
$\rho 15$	-0.15	-0.167	0.019	0.026	-0.211	-0.199	-0.167	-0.136	-0.123
$\rho 23$	-0.10	-0.148	0.019	0.026	-0.191	-0.178	-0.148	-0.116	-0.104
$\rho 24$	-0.25	-0.219	0.019	0.025	-0.262	-0.250	-0.219	-0.188	-0.174
$\rho 25$	-0.17	-0.082	0.024	0.029	-0.134	-0.119	-0.082	-0.045	-0.028
$\rho 34$	-0.12	-0.159	0.018	0.024	-0.201	-0.189	-0.159	-0.128	-0.116
$\rho 35$	-0.10	-0.142	0.019	0.026	-0.186	-0.174	-0.142	-0.111	-0.098
$\rho 45$	-0.18	-0.152	0.022	0.028	-0.200	-0.186	-0.152	-0.117	-0.101

Notes: Results based on 10,000 iterations, discarding the initial 100. Local coherency checked. SD = standard deviation; IQR = interquartile range; P1, P5, P95, P99 = the 1st, 5th, 95th, and 99th percentiles, respectively.

5. Conclusion

The proper estimation technique when estimating complete demand systems using disaggregated data at the individual-, household-, or firm-level containing many non-consumed goods remains an important, but unanswered, question. Two problems arise in the estimation of such models. Firstly, the model represents an endogenous switching regimes model, involving the evaluation of high-dimensional probability integrals. Secondly, the problem of statistical coherency must be addressed. The curvature conditions imposed by demand theory, concavity of the cost or utility function, are sufficient, but may be unduly restrictive in many situations.

As first articulated in Pitt and Millimet (1999), the use of the Gibbs sampling algorithm, along with data augmentation, solves both of these issues. Given the simple latent structure of the demand system, augmenting the data removes the need for integration. In addition, use of rejection sampling within each loop of the Gibbs sampler allows one to impose coherency only where desired: at each data point, within a defined space surrounding the observed data points, or globally. Previous attempts to ensure coherency in models with corner solutions have not been successful. Either more restrictive functional forms have been utilized, such that coherency is guaranteed globally, or global coherency has been imposed on more flexible functional forms,

destroying their flexibility, or less than fully structural demand systems have been estimated.

Here, we build on Pitt and Millimet (1999), utilizing Gibbs' sampling and data augmentation to estimate translog indirect utility and cost functions in a computationally convenient way. Results using several simulated data sets confirm not only the accuracy of the Gibbs sampler estimates, but also the importance of addressing the problem of coherency. In all of the non-linear simulated data sets, the parameter vector frequently enters the space of incoherent values even though the 'true' data-generating process is globally coherent. Moreover, between roughly 40%–50% of parameter draws that satisfy local coherency, fail to satisfy global coherency. Thus, rejecting on the basis of global coherency significantly restricts the acceptable parameter space relative to the model imposing coherency locally.

Acknowledgements

The authors thank Justin Tobias, Juan Carlos Escanciano, Konstantin Tyurin, and seminar participants at SMU. The authors are grateful for comments from an anonymous referee.

References

- Arndt, C., 1999. Demand for herbicide in corn: An entropy approach using micro-level data. *Journal of Agricultural and Resource Economics* 24, 204–221.

- Barnett, A.B., Geweke, J., Wolf, M., 1991. Semiparametric Bayesian estimation of the asymptotically ideal production model. *Journal of Econometrics* 49, 5–50.
- Chakir, R., Bousquet, A., Ladoux, N., 2004. Modeling corner solutions with panel data: Application to the industrial energy demand in France. *Empirical Economics* 29, 193–208.
- Christensen, L., Jorgensen, D., Lau, L.J., 1975. Transcendental logarithmic utility functions. *American Economic Review* 53, 367–383.
- Diewert, W.E., Wales, T.J., 1987. Flexible functional forms and global curvature conditions. *Econometrica* 55, 43–68.
- Dong, D., Gould, B.W., Kaiser, H.M., 2004. Food demand in Mexico: An application of the Amemiya–Tobin approach to the estimation of a censored food system. *American Journal of Agricultural Economics* 86, 1094–1107.
- Gallant, A.R., Golub, E.H., 1984. Imposing curvature restrictions on flexible functional forms. *Journal of Econometrics* 26, 295–321.
- Golan, A., Perloff, J.M., Shen, E.Z., 2001. Estimating a demand system with nonnegativity constraints: Mexican meat demand. *Review of Economics and Statistics* 83, 541–550.
- Gourieroux, C., Laffont, J., Montfort, A., 1980. Coherency conditions in simultaneous linear equation models with endogenous switching regimes. *Econometrica* 48, 675–695.
- Hausman, J., 1985. The econometrics of nonlinear budget sets. *Econometrica* 53, 1255–1282.
- Jorgenson, D.W., Fraumani, B.M., 1981. Relative prices and technical change. In: Berndt, E., Field, B. (Eds.), *Modelling and Measuring Natural Resource Substitution*. MIT Press, Cambridge, MA.
- Kao, C., Lee, L.-F., Pitt, M.M., 2001. Simulated maximum likelihood estimation of the linear expenditure system with binding non-negativity constraints. *Annals of Economics and Finance* 203–223.
- Lau, L.J., 1978. Testing and imposing monotonicity, convexity, and Quasi-convexity constraints. In: Fuss, M., McFadden, D. (Eds.), *Production Economics: A Dual Approach to Theory and Applications*, vol. 1. North-Holland, Amsterdam, pp. 409–453.
- Lee, L.F., Pitt, M.M., 1986. Microeconomic demand systems with binding nonnegativity constraints: The dual approach. *Econometrica* 54, 1237–1243.
- Lee, L.F., Pitt, M.M., 1987. Microeconomic models of rationing, imperfect markets, and non-negativity constraints. *Journal of Econometrics* 36, 89–110.
- Lewbel, A., 1997. Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica* 65, 1201–1214.
- MaCurdy, T., Green, D., Paarsch, H., 1990. Assessing empirical approaches for analyzing taxes and labor supply. *Journal of Human Resources* 25, 415–490.
- Meyerhoefer, C.D., Ramney, C.K., Sahn, D.E., 2005. Consistent estimation of censored demand systems using panel data. *American Journal of Agricultural Economics* 87, 660–672.
- Neary, J.P., Roberts, K.W.S., 1980. The theory of household behavior under rationing. *European Economic Review* 13, 25–42.
- O'Donnell, C.J., Shumway, C.R., Ball, V.E., 1999. Input demands and inefficiency in U.S. agriculture. *American Journal of Agricultural Economics* 81, 865–880.
- Pitt, M.M., Millimet, D.L., 1999. Estimation of Coherent Demand Systems with Many Binding Non-Negativity Constraints, Brown University Department of Economics WP 1999-4.
- Ransom, M.R., 1987. A comment on consumer demand systems with binding non-negativity constraints. *Journal of Econometrics* 34, 355–359.
- Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* 82, 528–549.
- Terrell, D., 1996. Incorporating monotonicity and concavity conditions in flexible functional forms. *Journal of Applied Econometrics* 11, 179–194.
- van Soest, A., Kapteyn, A., Kooreman, P., 1993. Coherency and regularity of demand systems with equality and inequality constraints. *Journal of Econometrics* 57, 161–188.
- van Soest, A., Kooreman, P., 1990. Coherency of the indirect translog demand system with binding nonnegativity constraints. *Journal of Econometrics* 44, 391–400.
- Wales, T., Woodland, A., 1983. Estimation of consumer demand equations with binding non-negativity constraints. *Journal of Econometrics* 21, 263–285.
- Yen, S.T., Lin, B.-H., Smallwood, D.M., 2003. Quasi- and simulated-likelihood approaches to censored demand systems: Food consumption by food stamp recipients in the United States. *American Journal of Agricultural Economics* 85, 458–478.