The Welfare Effects of Bundling in Multichannel Television Markets

Gregory S. Crawford and Ali Yurukoglu

Data and Code Readme File

1 Introduction

This file documents how to produce results comparable to those we obtained in our paper, "The Welfare Effects of Bundling in Multi-Channel Television Markets." They do not reproduce our exact results due to licensing restrictions that prevent us from sharing much of the raw data we use in the paper. More details are provided below.

There are two parts to this readme document:

1. Data Sources

   The first part of the second section of this readme file describes the six primary datasets used in our paper and how to obtain access to them. All but one are proprietary and require either a license agreement from the data provider and/or significant effort to obtain the data by hand from published sources.

   These datasets were merged using Stata and output into comma-delimited (CSV) files for use by the estimation code. The Stata code used to do this is enormous. As this code is unlikely to be useful absent the raw data, we have chosen not to provide it with the estimation code. Anyone interested in it, however, should contact us and we will provide any or all parts that may be useful to them.

   While we cannot provide the true CSV files used in our estimation, we provide comparable "Fake Data" that shares the same structure as our true data and allows the estimation code to run to completion. The second part of the second section of this readme file describes each of these data files.

2. Estimation Code

   The third section of this readme file provides a list of Matlab and Fortran 90 program files used in the estimation of the paper. These program files run on the Fake Data to produce results comparable in structure to those we obtained in our paper. We provide the calling structure of these program files and briefly describe what each does below.

   The Matlab code uses both optimization routines from TOMLAB and Matlab "mex" files generated using Fortran 90. The former may be obtained for a 21-day trial period from URL: `http://tomopt.com/tomlab/`, after which a license must be purchased. The latter requires

a Fortran 90 compiler compatible with Matlab, of which there are many. See the Matlab documentation for help if needed.

2 Data

2.1 Data Sources

We used six data sources in our paper, five of which are proprietary. We describe how to access each of these below.

1. Cable System (Factbook) Data

   The Television and Cable Factbook Electronic Edition provides data at the local cable market level on the composition of cable television bundles, their prices and market shares, cable system ownership, and other system characteristics. Factbook data is compiled and licensed by Warren Communications, LLC. We use Factbook data from 1997-2007.

   Factbook data is made available in two forms: (1) electronically and (2) in annual printed publications. The electronic version is a snapshot of the information maintained in Warren Communications database at the time of access. Unfortunately, Warren does not keep or license historical information about cable systems. The only way we know to access historical information about cable systems is to digitize the information contained in the annual printed Factbooks.

   Those interested in obtaining access to current Factbook data may contact their representative below.

   Brook Mowry
   Account Manager
   Warren Communications News
   2115 Ward Court, NW
   Washington, DC 20037
   202-872-9202, ext. 206
   bmowry@warren-news.com
   www.warren-news.com

2. Satellite Data

   Information about the composition of (national) satellite service bundles and their prices were collected by hand using information published by individual satellite companies (DirecTV and Dish Network). These were accessed using the Wayback Machine at the Internet Archive (`http:\archive.org`).

   We then matched this to aggregate satellite cable market data data at the Designated Market Area (DMA) level using data on market shares provided on "Alternative Delivery Systems"

(ADSs) by DMA published by the Television Advertising Bureau (`http:\www.tvb.org`).[1] As of June 2011, TVB maintains ADS market shares by DMA back to February 2000. The ultimate source of this data is Nielsen Media Research and can surely also be obtained directly from them. We provide contact information for Nielsen Media Research below.

3. Cable Channel Data

   Information about national (across-system) average license fees paid to individual cable networks was obtained from the 2006 Edition of the Economics of Basic Cable Network. This was licensed from Kagan Media Research which has since been purchased by SNL and re-named SNL Kagan. Current and historical information about cable network license fees are now available with an annual subscription to a variety of SNL Kagan media and communications databases.

   Those interested in obtaining access to SNL Kagan data may contact one of their sales representatives (of which there are many). We list one with whom we have worked below.

   Michael Bengel
   SNL Kagan
   Sales Executive
   212.542.8030
   `www.snl.com\media_comm`

4. Aggregate Viewing Data

   Current and historical information about average ratings (viewing) of individual television channels (or, indeed, individual television programs) across DMAs and/or time can be obtained from Nielsen Media Research.

   Those interested in obtaining access to Nielsen Media Research data may contact their representative listed below.

   Carly Litzenburger
   Account Executive, Media
   The Nielsen Company
   646-654-8366
   Carly.Litzenberger@nielsen.com

5. Household Viewing Data

   Current and historical information about the viewing of individual cable channels can be obtained from Mediamark Research and Intelligence (MRI).

[1] For example, see `http://www.tvb.org/planning_buying/184839/4729/72555`.

Those interested in obtaining access to MRI data may contact their representative below.

Adriane Heimann
Supervisor, Client Services
212.884.9283
Mediamark Research & Intelligence

6. Census Data

   Current and historical information about demographic information across space and time is available for free from the U.S. Census. We used information from the 2000 Census Summary Data File 3, available using the "Data Ferrett" tool at `http:\www.census.gov`. We linked this information to the cable systems in the Factbook by linking the community names served by each cable system to the zip codes served by that community and averaging across all the zip codes served by that system.

2.2 Structure of the "Fake Data"

The fake data are provided in a directory called "Fake Data". They are comparable in structure to the real data we used in estimation, but their contents have been randomized to protect the intellectual property of the rights holders.

They are input in the following order. The calling matlab (".m") file in which they are called is given at the heading of each list of data files.

From data_import.m (inside main.m)

1. "dma_market_and_bundle_id.csv": This file provides identifying information for each of the bundles in the estimation data. An observation is a Year-DMA-Market-Bundle. A market is an area served by a single cable system. Years vary from 1997 to 2007, DMAs from 1-210, markets within DMA according to how many cable markets (systems) there are in a DMA (max = 109), and bundles according to how many bundles the system offers (max = 4). The columns in the file are:
   - Year
   - DMA: Nielsen Designated Market Area
   - DMA_Market_ID: An identifier for each market within each DMA
   - DMA_Market_Count: The number of markets within that DMA
   - DMA_BundleID: An identifier for each bundle within each DMA
   - DMA_Bundle_Count: The number of bundles within that DMA
   - Bundle_Count: The number of bundles offered by the system in each market
   - Bundle_ID: An identifier for each of those bundles
2. "channels.csv": This file provides dummy variables indicating whether each of 375 channels is offered on the bundle in each row. The names of these 375 channels are available in the file options.m or in the Matlab .mat file "names.mat." Note the names listed are those we

used in the estimation, but the underlying data may not match that information due to the randomization process we used to create the fake data.

3. "bundle_dummies.csv": This file provides 88 dummy variables used in the estimation. The columns in the file are:

   - MSO Dummies (40): Dummy variables for each of 40 Multiple System Operators (cable system conglomerates) in the data.
   - Tier-of-service Dummies (5): Dummy variables for each tier of service in the data (Basic, Expanded, Digital, etc.), with Basic Service the excluded category.
   - Year Dummies (10): Dummy variables for years (1997-2997), with 1997 the excluded category.
   - Number-of-bundles-offered Dummies (3): Dummy variables for number of bundles offered (1-4), with 1 bundle the excluded category.
   - Year $\times$ Number-of-bundles-offered-dummies (30): All interactions of the two dummy variables above.

4. "prices.csv": This file provides the (fake) price (in year-2000 dollars) for each of the bundles in the data.

5. "dma_avg_prices.csv": This file provides the average price for cable service for all other systems in that DMA excluding the system in each row. This is used as a instrument for prices in the estimation.

6. "mshare.csv": This file provides the (fake) market share for each of the bundles in the data.

7. "homespassed.csv": This file provides the number of homes passed for each cable system in the data.

8. "relative_size.csv": This file provides the size of the system compared to the other systems in that DMA in that year.

9. "dma_dummies.csv'): This file provides a dummy variable for each of the DMAs in the data, with DMA 1 (New York) the excluded category.

10. "more_weight_in_viewership.csv": In order to use the Factbook and MRI viewership together, we needed to weight the systems in the Factbook. This file indicates which bundles on which systems should be weighted more heavily.

11. "demographics.csv": This file reports the average demographic variables for the zip codes served by the system in each row. The columns in the file contain information on:

    - Year
    - DMA
    - Family Status (2 fields): The % of households of size 2 or more and of size 3 or more
    - Income (8 fields): The % of households earning \$0-10, \$10-25, \$25-40, \$40-60, \$60-75, \$75-100, \$100-150, and \$150 thousand or more.
    - Urban/Rural (2 fields): The % of households that live in an urban (rural) area
    - Race/Ethnicity (4 fields): The % of households that are white/black/asian/hispanic

- Education (2 fields): The % of households that have a bachelors (higher) degree
- Age (5 fields): The % of households age 0-12, 13-18, 18-29, 30-39, and 40-59.

12. "homes_passed_share.csv": When using the Factbook and MRI viewership together, we weighted each Factbook observation by its size as measured by the number of homes passed. This was done for bundles/systems after 2000 as those were the years for which we have MRI data.

13. "satellite_penetration_yyyy.csv" (11 files, one each for 1997-2007): This file reports two types of market share information for each of the 210 DMAs. The columns in the file are:

    - Market share (within DMA) of satellite television services
    - Market share of the outside good (no television)

14. "directv.csv": This file contains information on the most popular DirecTV service between 1997 and 2007. For each year, it reports the price of that service and dummy variables indicating that service offered each of the 375 channels enumerated above.

15. "nielsen_ratings.csv": This file reports (fake) aggregate Nielsen ratings by DMA for each of 56 channels. A row is a dma-year-month. The columns in this file contain information on:

    - DMA Index
    - DMA Size
    - Month (February, May, July, November)
    - Channels (375)
    - Average demographics in that DMA (urban %; family %; avg income (00,000s); black %; hispanic %; asian %; college %; age (00s))
    - Year

16. "mediamark_full.csv": This file reports (fake) household-level data from MRI at the level of the year-channel. A row in the observation is an MRI household in a given "wave" of data (from which one can infer the year). The columns in this file contain information on:

    - Household ID: An identifier for each MRI household
    - Wave: The MRI wave (an identifier MRI used to indicate the time period the data was collected)
    - Channels (375): The hours watched of that channel by that household
    - Various demographic information about that household (household size; income; whether white/black/asian/hispanic; whether college/higher degree; age of respondent; whether subscribe to cable/digital cable/DirecTV/Dish)

From create_households_from_individual_data.m (inside data_import.m)

1. "census.csv": This file contains raw census microdata. A row is a household. The columns in this file contain information on:

    - Household ID
    - Household size
    - Age of each person in household (up to 8)

- Sex of each person in household
- Race of each person in household
- Marital status of each person in household
- Education of each person in household (5 standard Census categories)
- Number of children
- Household income
- Whether the household owns or rents

From cost_estimation_main.m

1. "msoid.csv": This file contains an identifier for the MSO that owns the cable system providing the bundle offered in that market in that year.
2. "2006_vertical_integration_table.csv": This file contains information about the ownership share of each of our 375 channels (rows) across each of 8 MSOs (columns) that have ownership shares in any cable channel in 2006. The 8 MSOs are Comcast, Time Warner, Cox, Charter, Cablevision, Other, DirecTV, and Dish.
3. "pre_kagan_costs.csv": This file contains (fake) across-cable-system average affiliate fees paid by cable systems to each of the 375 channels in our analysis.

From bargaining_main.m

1. "channel_ad_revenue.csv": The (fake) annual per-subscriber advertising revenue for each channel in the data.

From compute_elasticities.m

1. "tierno.csv": This file contains the tier of service associated with each bundle.

3 Matlab Code

We estimate our model in three separate steps:

1. Demand estimation
2. Cost estimation
3. Bargaining estimation and counterfactual simulation

We describe the matlab files called by each of these steps in what follows

3.1 Demand Estimation

The structure of the matlab and fortran files for the demand estimation is as follows:

- main.m: The main demand estimation file
  - `C:\tomlab\startup.m`: Enables the use of TOMLAB routines

- options.m: Set various data and estimation options
- data_import.m: Import the various datasets needed for demand estimation.
  - create_households_from_individual_data.m: The MRI data are for individuals. This file creates MRI households by adding up the viewing of MRI individuals in combinations similar to the kinds of households seen in Census data.
- set_random_draws.m: This file creates one set of random households for calculating market shares and another for calculating viewing shares.
  - latin_random.m: Draw random variables from a Latin Random Hypercube
  - (latin_random.m itself calls i4_uniform.m, r8_uniform_01.m, perm_random.m)
- create_dummy_utility_parameters.m: Breaks down full parameter vector into relevant components
- full_optimization.m: Runs the demand estimation. Prior to full_optimization being called, the code in main.m provides a number of options for the user to set about which sets of demand parameters should be estimated (all/no copula) and how that estimation should be done (univariate/channel-by-channel/joint). Univariate here mean one parameter at a time, even within a channel. Channel-by-channel means jointly estimating all parameters associated with that channel. Joint means estimating all parameters associated with all channels.
  - full_objective_function.m: Sets up the demand moments and objective function for the demand estimation.
    - assemble_markets_fortran_mex.f90 (fortran file): For given parameters and draws, this file computes all the BLP interactions
    - delta_contraction.f90 (fortran file): This file does the Berry/BLP contraction: it determines mean utility levels for each bundle in each market given all non-linear parameters.
    - market_shares.m: This file calculates market shares by aggregating over the bundle random households
    - group_stats.f90 (fortran file): This file computes viewership by demographic by aggregating over the viewership random households
  - fminsearchcon.m: Non-derivative constrained optimization algorithm
  - compute_eigenvalues.m: Code to make sure the covariance matrix of random tastes for channels, Sigma, is positive semi-definite
- standard_errors_demand.m: Calculate standard errors for demand estimates
  - d_full_objective_function.m: Calculate the derivative of the objective function
  - full_objective_function_for_S.m: As above for full_objective_function, but with bootstrap weights

3.2 Cost Estimation

The structure of the matlab and fortran files for the cost estimation is as follows:

- cost_estimation_main.m: The main cost estimation file
  - cost_options.m: Set options for the cost estimation
  - data_import.m: (As above)

- set_random_draws.m: (As above)
- cost_preliminaries.m:
  - assemble_markets_fortran_mex.f90 (fortran file): (As above)
  - create_DELTA_matrix.f90 (fortran): The "Delta" matrix is as in Nevo's Practitioner's Guide (JEMS). Used to solve for optimal prices and thus marginal costs.
- cost_objective_function.m: Sets up the cost moments and objective function for the cost estimation

3.3 Bargaining Parameter Estimation and Counterfactual Simulation

The structure of the matlab and fortran files for the bargaining parameter estimation and counterfactual simulation is as follows:

bargaining_main.m: The main bargaining parameter estimation file. At the top of bargaining_main.m, the user is able to set options about whether that run of the code should include estimation and/or counterfactual simulation, and what, if any, counterfactual regulatory environment (Baseline bundling/à la carte/bundle-size pricing/theme tiers) should be considered.

- cost_options.m: (As above)
- input_cost_specification.m: Estimates of input costs as starting values
- set_random_draws_single_market.m: As in set_random_draws.m, but for all US rather than market by market
- sample_wtp.m: Draws a number of consumers and evaluates their WTP
- bargaining_set_basic_parameters.m: Calibrates the size of the logit error for the synthetic Big/Small Cable markets
- uni_bargaining_estimation_objective_function.m: Sets up bargaining objective function for bargaining parameters estimation.
  - (Similar to bargaining_estimation_objective_function.m. The only difference is that "uni_" estimates parameters channel-by-channel instead of all at once)
- bargaining_equilibrium.m: Computes equilibrium of the bargaining game
  - bargaining_disagreement_payoffs.m: This file and the next (bargaining_downstream_equilibrium) computes disagreement payoffs if bargains can't be reached for (1) baseline bundle, (2) bundle-sized pricing (BSP), and (3) à la carte (ALC).
    - bargaining_downstream_equilibrium.m: As immediately above. There are 3 options for this code:
      1. Baseline: Computes downstream prices, up-down profit split, and more under baseline bundles. Also one ALC file (see below).
         - bargaining_downstream_firm_profits.m: As immediately above.
         - bargaining_alc_time_watched_mex.f90 (fortran): Computes time watching channels under ALC (at existing bundle costs)

2. BSP: Computes downstream prices, up-down profit split, and more under BSP.
   - bargaining_bsp_preliminaries.m: Compute some BSP inputs (e.g. 10 favorite channels)
   - bargaining_bsp_downstream_firm_profits_mex.f90 (fortran): Computes downstream prices and profits under BSP
   - bargaining_bsp_time_watched_mex.f90 (fortran): Computes time watching channels under BSP
3. ALC: Computes downstream prices, up-down profit split, and more under À La Carte.
   - uni_bargaining_alc_downstream_firm_profits_mc_pricing.m: Computes downstream prices and profits under ALC assuming marginal cost pricing of channels downstream.
   - (Similar to bargaining_alc_downstream_firm_profits_mc_pricing.m. Only difference is that "uni" estimates channel by channel)
   - (Calls bargaining_alc_downstream_firm_profits_mc_pricing_channel_mex.f90 (fortran): Computes downstream prices and profits per channel under ALC)
   - bargaining_alc_downstream_firm_profits_channel_mex.f90 (fortran): *** Optional *** No longer used for the results in the paper due to assumption of specific pricing rule downstream. Included in case one wants to solve for optimal downstream ALC prices.
   - bargaining_alc_time_watched_channel_mex.f90 (fortran): Computes time watching channels by channel under ALC

* bargaining_downstream_equilibrium_tt.m: This file computes disagreement payoffs if bargain can't be reached for theme tiers (TT)
  - bargaining_tt_downstream_equilibrium_preliminary.m: Computes some preliminary inputs for theme tiers
  - (Calls uni_bargaining_tt_downstream_firm_profits_fast.m: Computes downstream prices and profits under theme tiers)
  - (Which itself calls bargaining_tt_downstream_firm_profits_fast_mex.f90 (fortran): As immediately above, except "uni" estimates channel channel by channel.)
  - bargaining_tt_preliminary.m : Defines the channels in the theme tiers.
  - bargaining_tt_downstream_firm_profits_mex.f90 (fortran): (As above)
  - bargaining_tt_time_watched.f90 (fortran): Computes time watching channels under theme tiers

- bargaining_nash_product.m: Calculates Nash Products given profits and threat points upstream and down
  - bargaining_downstream_equilibrium.m: (As above)
  - bargaining_downstream_equilibrium_tt.m: (As above)

3.4 Other Standalone Matlab programs

- compute_elasticities.m: Computes the matrix of own- and cross-price elasticities
  - market_shares.m: (As above)

  - create_DELTA_matrix.m: (As above)

- bargaining_welfare_tables.m: Construct tables of relevant economic outcomes
  - bargaining_welfare_measures.m: Calculate economic outcomes (prices, shares, welfare, etc.) for a given environment (bundling, alc, bsp, tt) and parameter values

4 Conclusion

If anyone has any questions or comments, they should feel free to contact either of us using our contact information below.

Gregory S. Crawford
Department of Economics
University of Warwick
Coventry CV4 7AL, UK
Ph: +44 (0)2476 523470
Email: crawford@warwick.ac.uk

and/or

Ali Yurukoglu
Graduate School of Business
Stanford University
Stanford, CA 94305
Ph: +1 (650) 721 1293
Email: Yurukoglu_Ali@gsb.stanford.edu