

United States  
Department  
of Agriculture



Economic  
Research  
Service

# On the Accuracy of Nielsen Homescan Data

Economic  
Research  
Report  
Number 69

Liran Einav  
Ephraim Leibtag  
Aviv Nevo

December 2008



[www.ers.usda.gov](http://www.ers.usda.gov)

## Visit Our Food Sector Briefing Room To Learn More!

ERS develops data describing the economic health and dynamics of the food sector. ERS research analyzes the causes of structural change in the food sector and its impact on food markets and consumers.

[www.ers.usda.gov/Browse/FoodSector/](http://www.ers.usda.gov/Browse/FoodSector/)

## National Agricultural Library Cataloging Record:

Einav, Liran

On the accuracy of Nielsen Homescan data.  
(Economic research report (United States. Dept. of Agriculture. Economic Research Service) ; no. 69)  
1. Grocery shopping—United States. 2. Consumers—Research—Validity—United States.  
3. Groceries—Prices—United States. I. Leibtag, Ephraim. II. Nevo, Aviv. III. Title.

TX356

Photo credit: Shutterstock

The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and, where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or a part of an individual's income is derived from any public assistance program. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD).

To file a complaint of discrimination write to USDA, Director, Office of Civil Rights, 1400 Independence Avenue, S.W., Washington, D.C. 20250-9410 or call (800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA is an equal opportunity provider and employer.



United States  
Department  
of Agriculture

Economic  
Research Report  
Number 69

December 2008



A Report from the Economic Research Service

[www.ers.usda.gov](http://www.ers.usda.gov)

# On the Accuracy of Nielsen Homescan Data

**Liran Einav**

**Ephraim Leibtag**

**Aviv Nevo**

## Abstract

Researchers use Nielsen Homescan data, which provide detailed food-purchase information from a panel of U.S. households, to address a variety of important research topics. However, some question the credibility of the data since the data are self-recorded and the recording process is time-consuming. Matching purchase records from 2004 Homescan data with data obtained from a large grocery retailer, it is evident that quantities purchased are reported more accurately in Homescan than are prices. Many of the price differences may be driven by the way Nielsen imputes prices: when available, Nielsen uses store-level prices instead of the actual price paid by the household. There are also differences by household type in the tendency to make mistakes that are correlated with demographic variables. However, the fraction of variance explained by the documented recording errors is in line with other research data sets for which cross-validation studies have been conducted.

**Keywords:** Nielsen, Homescan, scanner data, validation study

## Acknowledgments

The authors would like to thank Andrea Pozzi and Chris Taylor for their research assistance and Oral Capps, Andrea Carlson, Helen Jensen, and Jessica Todd for their helpful reviewer comments. This research was funded by a cooperative agreement between the ERS and Northwestern University.

## About the Authors

Liran Einav is an associate professor of economics at Stanford University, Ephraim Leibtag is an economist with ERS, and Aviv Nevo is a professor of economics at Northwestern University. Einav and Nevo are affiliated with the National Bureau of Economic Research (NBER), Cambridge, MA.

**Contents**

**Summary** . . . . . iii

**Introduction** . . . . . 1

**Data Sources** . . . . . 3

    Homescan . . . . . 3

    Retailer’s data . . . . . 4

**Data Construction** . . . . . 5

    First Step . . . . . 5

    Second Step . . . . . 6

    Summary . . . . . 7

**Record-Matching Strategy** . . . . . 8

    Results . . . . . 9

    Trip Information . . . . . 10

    Product Information . . . . . 15

    Price and Quantity Information . . . . . 18

**Do the Differences Matter?** . . . . . 24

**Conclusions** . . . . . 26

**References** . . . . . 28

**Recommended citation format for this publication:**  
Einav, Liran, Ephraim Leibtag, and Aviv Nevo. *On the Accuracy of Nielsen Homescan Data*. ERR-69, U.S. Dept. of Agriculture, Econ. Res. Serv. December 2008.

## Summary

Researchers use Nielsen Homescan data, which provide detailed food-purchase information from a panel of U.S. households, to study the dynamics of retail food markets.

### *What Is the Issue?*

Some questions have been raised regarding the credibility of the Nielsen Homescan data because the data are self-recorded and the recording process is time-consuming. Given the time commitment, households who agree to participate in the sample might not be representative of the U.S. population as a whole, and those who agree to participate may not record their purchases accurately.

### *What Did the Study Find?*

The analysis conducted in this report suggests that the Homescan data contain recording errors in several dimensions, but that the overall accuracy of self-reported data by Homescan panelists seems to be in line with other commonly used (government-collected) economic data sets.

For approximately 20 percent of food-shopping trips recorded in the Nielsen Homescan data, there was no corresponding transaction in the retailer's data, suggesting that either the store or date information was recorded with error. Using the retailer's loyalty card information, the study finds some shopping trips that did not match up with Nielsen Homescan data, implying that households did not record all of their trips in their Homescan records.

For the trips that did match up, roughly 20 percent of the items purchased were not recorded. For those items that were recorded, quantity was reported fairly accurately: 94 percent of the quantity information matched in the two data sets. The match for prices was lower: in almost half of the cases, the two data sets did not agree. However, much of this difference can be attributed to transactions that involved promotional or other temporary sale prices in either the Nielsen Homescan data or the retailer's data.

Nielsen's practice of using store-level data as an estimate of what households actually paid poses a challenge when those stores have multiple possible prices in a given time period due to loyalty card or other shopper-specific price promotions. Indeed, for prices that involve no promotion or temporary price reduction, there are recording errors in only about 17 percent of the cases. Therefore, much of the overall price difference is likely caused by the way Nielsen imputes prices and not by recording errors by the panelists. Mismatched prices would most likely be less of a problem for stores that only have one price per product in a given week, so that the results highlight the importance of store pricing practices in food price analysis.

The study also compares the recording errors to errors in other commonly used economic data sets, and finds that errors in Homescan are of the same order of magnitude, for example, as reporting errors in earnings and employment-status data.

### *How Was the Study Conducted?*

Homescan records contain all products purchased by a household on a particular day in a particular store, as they were scanned by the consumer. The study compared these records to data obtained from a single retailer. The retailer's data contain the products purchased in each of the transactions at the same store and day reported by the household, as recorded by the cashier. Using data from trips made during 2004, the records from both data sets were matched. The matched transactions were compared and contrasted, and differences in various dimensions were recorded. In order to study the impact the recording errors might make in an applied study, the price paid was regressed on household characteristics in both data sets to see if the results differ.