Is IT Enough? Evidence from a Natural Experiment in India's Agriculture Markets

Chris Parker
Smeal College of Business, Pennsylvania State University, University Park, PA 16802
chris.parker@psu.edu

Kamalini Ramdas, Nicos Savva
London Business School, Regent's Park, London NW1 4SA, United Kingdom,
kramdas@london.edu, nsavva@london.edu

Access to information and communication technologies (ICTs) such as mobile phone networks is widely known to improve market efficiency. In this paper, we examine whether access to timely and accurate information provided through ICT applications has any additional impact. Using a detailed dataset from Reuters Market Light (RML), a text message service in India that provides daily price information to market participants, we find that this information reduces the geographic price dispersion of crops in rural communities by an average of 12%, over and above access to mobile phone technology and other means of communication. To identify the effect of information on price dispersion we exploit a natural experiment where bulk text messages were banned unexpectedly across India for 12 days in 2010. We find that besides reducing geographic price dispersion, RML also increases the rate at which prices converge across India over time. We discuss the implications of this for development organizations and information providers.

Key words: price dispersion, information and communication technology, natural experiment, agricultural supply chains

1. Introduction

New information and communication technologies (ICTs) often create new ways to gather the information necessary to make economic decisions (Aker and Mbiti 2010, Mittal et al. 2010). For example, farmers and fishermen in rural areas of the developing world are using mobile phones to access information on the price of agricultural commodities in local markets (see review by Jensen 2010). By reducing the cost of access to such information, new ICTs enable farmers to make better decisions about where to sell their produce – shifting supply from low- to high-price markets, as well as when to sell it – delaying or bringing forward the harvesting or selling of crops to exploit price variation over time. Reducing the cost of acquiring such information should, in theory, yield a more precise matching of supply and demand and therefore result in more efficient markets and less variation in prices. Indeed, two recent studies report that the improved information flow associated with the introduction of mobile phone coverage caused a permanent decrease in the geographic price dispersion of fish in Kerala, India (Jensen 2007), and grain in Niger (Aker 2010).

Research to date implicitly assumes that the primary barrier to information acquisition is the prohibitive cost of communication, i.e., once communication costs are reduced by new ICTs such as mobile phones, information becomes readily available, resulting in a permanent reduction in price dispersion. While being able to communicate cheaply is clearly necessary for information acquisition, is it sufficient? After all, farmers, who may increasingly have easy access to affordable mobile phone networks, often do not have access to informed and unbiased parties from whom they can obtain timely and accurate information. The primary goal of this study is to investigate empirically whether the existence of a third-party information provider that farmers and other market participants can rely on for timely and accurate information transmitted through mobile phones has an impact on the matching of supply and demand of agricultural commodities, over and above the now widely recognized impact of having access to mobile phone infrastructure.

This is an important question because substantial resources are being invested in improving the efficiency of agricultural supply chains in the developing world; e.g., between 2003 and 2010 the World Bank invested $4.2 billion in the developing world's ICT infrastructure (World Bank 2011). Most of this funding has been used to improve access to ICTs, which remains limited particularly in rural areas, such as rural India, where mobile phone penetration is estimated to be 38% (IAMAI 2012). To foster the welfare improvements that have been shown to accrue from a reduction in price dispersion (see Jensen 2007), should governments and funding agencies continue to invest heavily in improving ICT infrastructure in developing nations, or should they complement this investment with direct or indirect support for third-party ICT applications that provide reliable and up-to-date information to market participants?

To answer this question we use a detailed dataset from Reuters Market Light (RML), a commercial third-party information provider. RML provides information on the price of agricultural commodities in India via mobile phone text messages sent daily to its paying subscribers. We use actual subscriber data to show that by using this information RML subscribers can make better-informed decisions, e.g., a farmer willing to travel up to 100 miles to trade will receive a 6.5% higher price on average than he would at his closest market and he will only actually have to travel an additional 20.4 miles on average. In exploiting this information subscribers enable a better match between supply and demand, which we would expect to reduce geographic and temporal price dispersion. However, measuring the impact of RML-provided information on geographic price dispersion is an econometrically challenging problem due to endogeneity: if the benefit of having access to accurate price information is greatest in areas with chronically high price dispersion, one might expect these to be the areas where RML has the most subscribers. By simply comparing price dispersion in areas where RML has a relatively high penetration to that in areas where it does not, one would likely underestimate the reduction in price dispersion due to RML, or indeed find that RML is associated with higher price dispersion. One possible solution to this potential endogeneity problem is to perform a randomized controlled trial. However, this would be expensive and disruptive to organize. Instead, we exploit a natural experiment to overcome this problem.

On September 23, 2010, RML experienced a major service disruption when bulk text messages were unexpectedly banned throughout India in advance of a high court verdict that was to be announced on September 24, 2010. The case in question was a land title dispute for a site in the city of Ayodhya in Uttar Pradesh, India, which has religious significance for both India's Hindus and Muslims. Deadly protests and political unrest surround the history of this site. To reduce the likelihood of rumors spreading via text message and of technology-coordinated riots, the government directed telecommunication providers to immediately disable bulk text messages throughout the country on the evening of September 22, 2010. The ban was lifted on the night of October 4, 2010 and operations at RML returned to normal on October 5, 2010. For the 12 days of the ban, RML collected price information without being able to transmit it.

The ban resulted in an unexpected, exogenous shock to the information service provided by RML. All other sources of information retrieval, including visiting or calling markets, sending personal text messages, asking friends, family, and fellow market participants, or even checking prices through the Internet or Internet-enabled kiosks (see Goyal (2010)), were unaffected by the ban. This provides an almost ideal natural experiment that alleviates endogeneity concerns.

To measure the impact of the information provided by RML on geographic price dispersion we estimate a difference-in-differences model. We use fixed effects to control for crop, market and temporal heterogeneity, as well as a two-way robust error structure to allow for arbitrary error correlation within each crop over time and across crops on each day. We find that the average geographic price dispersion, measured by the coefficient of variation of the price of a crop in all markets open in a state on any given day increased by 12% on average during the ban in areas where RML had a relatively high penetration, over and above the change in price dispersion in areas where penetration was relatively low. Price dispersion differences between areas with relatively high and relatively low penetration returned to the pre-ban levels when the ban was lifted. Although not directly comparable due to the different context, the magnitude of the reduction is similar to that attributed to the reduction caused by access to ICTs in the literature (cf. Aker 2008).

To verify the main result, we investigate a number of alternative explanations and conduct a series of robustness checks. We find no evidence to suggest that the events surrounding the ban had a direct effect on price dispersion other than through the information provided by RML, i.e., we find no difference in either the average volume transacted or the coefficient of variation of the volume transacted across markets, suggesting that the increase in price dispersion is not due to market participants ceasing to trade or shifting their trade systematically (as opposed to randomly) during the ban. Furthermore, we use two additional measures to quantify geographic price dispersion and three measures to quantify RML penetration; allow for crop heterogeneity using a random coefficient specification; do a placebo test by repeating the analysis using data from 2009 when the RML service was not disrupted; use a density-based clustering algorithm to define alternative market groupings based on geographic proximity rather than administrative state boundaries; estimate the models under alternative assumptions on the structure of the variance–covariance matrix; vary the time-window used in the analysis; and change the filters applied to the data in constructing the panel. All tests yield results that are consistent with the main conclusion.

We complement this investigation by examining how the information provided by RML affects temporal price dispersion. We hypothesize that when the price of a crop in any market deviates substantially from the average price, market participants, to the extent that they are aware of the price deviation, will adjust their selling and purchasing decisions by shifting supply (demand) away from markets that have a lower (higher) than average price. As a consequence, random deviations from the average price should not be persistent. One mechanism by which market participants become informed about prices in nearby markets is through the information provided by RML. We therefore formulate and estimate a dynamic model of temporal price deviations and show that during the ban, price deviations from the average price are more persistent, i.e., fluctuations from the mean take 18.3% longer to attenuate, than before or after the ban in those markets where RML has a relatively high penetration relative to those in which it does not. Thus, we show that the information RML provides is one mechanism for promoting the "law of one price" in the agricultural markets of rural India.

We conclude the paper by discussing practical implications of our analysis, as well as outlining its main limitations.

2. Literature Review

The "law of one price" predicts that prices of homogeneous goods exchanged simultaneously in open markets at different locations should not differ by more than transportation costs (Isard 1977). However, many studies empirically document the existence of price dispersion, while others provide analytical models that identify sufficient conditions, e.g., search costs, under which price dispersion is an equilibrium outcome. Baye et al. (2006) provide an excellent review of this literature.

This study complements a subset of this literature that investigates the impact of new ICTs on price dispersion in general and on the price dispersion of agricultural commodities in the developing world in particular. This line of research presents the argument that successive generations of ICTs, including the telegraph, telephone lines, and, more recently, mobile phones and the Internet, have transformed markets by reducing search costs (Malone et al. 1987, Bakos 1991, Brynjolfsson and Smith 2000, Tang et al. 2010, Overby and Forman 2015, Ghose and Yao 2011). The literature examines whether prices, and more importantly for our purposes, price dispersion, are affected as newer technologies are introduced. While the general finding is that the advent of a new ICT is associated with a reduction in price dispersion, the evidence suggests that new ICTs do not eliminate price dispersion altogether. For example, Gatti and Kattuman (2003) measure the online price dispersion of 31 products across seven European countries and find significant price dispersion both between countries and across product categories such as printers or computer games. Similarly, Baye et al. (2004) report significant price dispersion for consumer electronics sold online. Explanations put forward to explain persistent price dispersion include consumers' failure to compare prices even when search costs are small (Baye and Morgan 2001), bounded rationality (Baye et al. 2004), consumers' failure to internalize additional costs such as shipping fees (Einav et al. 2015), or firms' use of "obfuscation strategies" (Ellison and Ellison 2009).

In the specific context of agricultural goods, two studies have examined the impact of ICTs on price dispersion. Jensen (2007) examines fish markets in Kerala, India before and after the introduction of mobile phones. He reports a dramatic and permanent reduction in geographic price dispersion after the introduction of mobile phones, resulting in impressive welfare gains. Aker (2010) reports that the introduction of mobile phones in Niger resulted in a 10% decrease in price dispersion across grain markets. Both studies exploit the staggered introduction of mobile phones for empirical identification. Our research contributes to this stream by demonstrating that having access to an independent and reliable information provider can further reduce price dispersion, over and above the gains of having access to a mobile phone. The natural experiment identification strategy also provides further evidence of the impact of ICT use on geographic price dispersion in agricultural supply chains. In addition, the daily frequency of our data, coupled with the short-lived ban, allows us to investigate how price information affects the speed with which prices converge over time – unlike prior studies, which use weekly or monthly survey data.

Other studies investigate the impact of ICTs on prices received by farmers, rather than price dispersion, with mixed results. Goyal (2010) studies the impact of the introduction of Internet kiosks on soybean prices in Madhya Pradesh, India. The kiosks were introduced by the India Tobacco Company, a large conglomerate that also purchases soybeans. Besides offering information about the price of soya in local and wholesale markets through its "e-Choupal" program, it also offers farmers the option to sell directly to the company at a pre-agreed price and quality (Devalkar et al. 2011). Goyal (2010) reports a 1.9% increase in soybean prices for farmers. Svensson and Yanagizawa (2009) examine the impact of a radio program that provides Ugandan farmers with market data. They report a 15% increase in the price received by maize farmers with radios in areas where the service is offered. In contrast to these studies, Fafchamps and Minten (2012) conduct a randomized controlled trial where they work with RML to provide the RML service to farmers in only some villages in Maharashtra over a one-year period. While the farmers claimed to have used the RML information, there is no statistical evidence that RML had an impact on their income. Similarly, Futch and McIntosh (2009) report no effect on the income of farmers from the introduction of a village phone program in Rwanda. In contrast to these studies, we aim to assess aggregate market efficiency as measured by geographic price dispersion.

3. Empirical Setting

In this section we provide an introduction to India's agricultural sector and information on the RML service and the natural experiment. The descriptions are based on interviews with RML employees, agricultural experts, farmers, traders, and government officials, and secondary sources.

3.1. Indian Agricultural Markets and Supply Chain

The Indian agriculture sector contributes an estimated 18.1% of national GDP and employs an estimated 52% of the labor force (CIA 2012). An estimated 455.8 million Indians live in agricultural households, on incomes below the World Bank's official poverty line of $1.25 a day (Chen and Ravallion 2008, Table 6). Given the size and extreme poverty of this sector, even a modest efficiency improvement will have a dramatic effect on welfare and could go a long way towards reducing poverty: one of the United Nations' key Millennium Development Goals.[1]

To prevent the exploitation of farmers, agricultural produce markets in India are regulated through the Agricultural Produce Marketing (APM) Act, which requires states to regulate spot markets for agricultural produce, called "mandis." At the local level, this regulation is the responsibility of the Agricultural Produce Marketing Committees (APMCs), which decide where to establish markets and manage their day-to-day operations. In particular, APMCs are responsible for licensing and regulating all traders permitted to operate within these markets, as well as for setting the rules for and overseeing all market transactions (Thomas 2003).

APMC-regulated mandis are either auction or terminal markets. Auction markets are smaller. Here, farmers sell directly to traders through auctioneers, who are either APMC employees or commission agents paid by farmers. Auctioneers are responsible for conducting auctions, weighing produce, and coordinating payment and delivery. On arrival at a market, a farmer is assigned a number and waits in the corresponding parking spot. The auctioneer, along with an administrator, leads traders from spot to spot, holding auctions for goods as they progress. The goods are presented and bid on by the traders in an open-outcry, ascending first-price (English) auction. The produce is then weighed and the farmer is paid according to the price set in the auction. Produce purchased in auction markets is sorted, packaged, and shipped to either domestic markets, which are called

[1] See `http://www.un.org/millenniumgoals/` (accessed July 29, 2014) for more information.

terminal markets, or international destinations. In terminal markets, traders and large farmers sell substantial quantities to wholesalers, retailers, and, occasionally, end consumers. Each trader sets a price and buyers visit different traders' stalls to barter for goods.

3.2. Reuters Market Light (RML)

RML was founded in 2006 with seed funding from the Reuters Venture Board to offer highly personalized information to farmers and other market participants via daily text messages through the existing mobile phone infrastructure. For a subscription fee of approximately Rs. 80 ($2) per month, RML provides information on local market prices and volumes transacted, highly localized weather forecasts, and crop-specific advisory (such as which fertilizer to use or how deep to plant specific seed varieties), as well as national and international news stories related to agriculture (Markides 2009). The information is provided exclusively via mobile phone text messages (SMS).

Market participants, mainly farmers, subscribe to RML by purchasing a pre-paid three-, six-, or 12-month subscription card from RML's local distributors, who are mainly agricultural supplies stores. Activating the subscription involves calling RML's dedicated call center and selecting two crops and three markets for each crop for which they wish to receive price/volume information, the taluka (similar to a US county) for which they wish to receive weather forecasts, and one of nine languages (Bengali, English, Gujarati, Hindi, Kannada, Marathi, Punjabi, Tamil, or Telugu) for their text messages. The subscription is activated within two days.

To gather price and volume information, RML follows a careful process engineered to safeguard quality. Market reporters visit each market and record the high and low prices of the day and the total volume transacted.[2] In auction markets this is done by simply observing the daily auctions, while in terminal markets the market reporters visit each of the individual traders' stalls. They then confirm this data with the local APMC officials and transmit it to the central RML system via voice call or text message. An automated system then validates the information and flags obvious errors such as typos or unusual price patterns. It is then checked manually by a chief reporter, who also independently checks the prices for several markets and resolves any discrepancies with the relevant market reporter before submitting the final price and volume information to the system. This information is then relayed to subscribers via bulk text messages sent through a third-party SMS vendor. The bulk text messages are sent out in two groups depending on the crop and market involved: the first group is sent around noon and the second at the end of the day. By the time they receive this information, most subscribers will not be able to act on it until the following day; yet, as we show in §5 it is still actionable.

[2] We note that the high and low prices that RML reporters collect are the 95th and 5th percentiles of the distribution of prices. RML uses these prices as they are more stable than the actual maximum and minimum prices. Following advice given by RML managers, we use the high price throughout our analysis as this is the price for the highest quality crops and is less susceptible to fluctuations due to quality differences.

As of December 2012, RML provides coverage for more than 300 crops and 1,300 local markets across 13 states, with one million unique subscribers in 50,000 talukas. The crops RML covers include staples (e.g., multiple varieties of lentils, rice, onions, and potatoes), perishables (e.g., multiple varieties of apples, bananas, and tomatoes), and animal produce (such as milk and eggs). The company has received a number of awards, including the 2010 Rural Marketing Association of India award for the best Internet/SMS/mobile initiative and the 2010 World Business and Development Award (WBDA) conferred by the United Nations Development Programme, the International Chamber of Commerce, and the International Business Leaders Forum.[3]

3.3. The Natural Experiment

Ayodhya, a city in the northern Indian state of Uttar Pradesh, is at the center of a centuries-old religious dispute over a piece of land that both Hindus and Muslims consider sacred. Controversy and violence have surrounded this site since at least 1853. Riots related to the disagreement resulted in over 2,000 deaths in 1992 and between 1,000 and 2,000 further deaths in 2002. At the end of September, 2010 the Allahabad High Court was set to rule on the dispute to determine how the land would be distributed between the two communities. Fearing that text messages might be used to spread rumors about the verdict and coordinate deadly riots, the Ministry of Communications and Information Technology banned bulk text messages beginning on the evening of September 22, 2010, up until the night of October 4, 2010. During the ban all messages sent through third-party SMS vendors, such as the one used by RML, were disabled. The September 30, 2010, verdict granted two-thirds of the disputed land to Hindu communities and the rest to Muslim communities. In contrast to government expectations, no violent outbreaks or rioting were observed following the announcement.[4] RML continued to collect market information as usual during the ban, although, the ban made it impossible to transmit this information to subscribers.

4. Data and Variables

Our analysis makes use of two distinct databases, both provided by RML. The first database holds information on RML's 350,000 subscribers; it includes a subscriber identification number, the start and end date of the subscription, the taluka, district, and state the subscriber resides in, and the subscriber's choices of up to three markets for each of two crops. The second database contains trading information from 1,136 markets. For any market where a crop is traded on any given day, this database contains the name of the market, the market's district and state, and the volume transacted, as well as the high and low prices.

[3] Information based on the company's website `http://www.reutersmarketlight.com/` (accessed July 29, 2014).

[4] See `http://www.guardian.co.uk/commentisfree/belief/2010/sep/30/politics-not-faith-brings-violence` and `http://www.ndtv.com/article/india/ayodhya-verdict-temple-politics-in-uttar-pradesh-56379` (accessed July 29, 2014) for examples of news media highlighting the absence of violence following the verdict.

We complement these datasets by using Amazon Mechanical Turk, a micro-outsourcing service that allows users to pay individuals to answer typically short and repetitive questions, to identify the geolocations of each market and the taluka of each subscriber in our database. We provided data on the state, district, and name of each market/taluka to individuals based in India and instructed them to use online geolocation tools such as Google Earth to find and report the longitude and latitude of the location of each market. The input data for each market was sent to two distinct workers. If the locations provided were within 10 miles of each other, we took the mid-point as the actual location. Where the responses differed by more than this threshold, the input data were re-submitted to two new workers. If we did not get consensus between these two workers a research assistant found the location by consulting physical maps at the British Library in London. Finally, we also use data on farmer land holdings from the most recent Agricultural Census (2006) and from Farmer Interest Groups (FIGs), which are local cooperative organizations that represent the interests of a sub-population of farmers who match RML's subscriber demographics.

4.1. Unit of Analysis

To investigate the impact of the RML information on geographic price dispersion we need to define an "area" in which we measure geographic price dispersion. Effectively, we assume that markets located within a given area affect each other, perhaps because some market participants are able to shift supply or demand from one market to the other, and are subject to similar exogenous shocks, such as weather. Conversely, markets in different areas are assumed to be sufficiently different or far away to preclude any strong co-movement of prices. One possibility would be to treat India as a single area. However, given the vast size of the country, which spans 1,997 miles (3,214 km) north–south and 1,822 miles (2,933 km) east–west, it is more reasonable to define smaller and more homogenous areas within which to examine price dispersion. As the trading of agricultural goods in India is heavily regulated by State Agriculture Marketing Boards, whose practices differ by state, one natural starting point for our analysis is to compare geographic price dispersion within state boundaries. Furthermore, since India is divided into 28 states largely on a linguistic basis, in addition to differences in regulation, linguistic and cultural differences are arguably more pronounced across state boundaries. In §6.4 we show that the results are not affected if we extend our analysis to arbitrarily-shaped "areas" generated using a density-based clustering algorithm.

On any given day, there are a number of markets open in each state trading a variety of crops. We call each individual market trading a specific crop a crop–market and collectively call the group of all open markets within a state trading a specific crop a crop–market-group. In §6 (§7), where we measure the impact of RML on geographic price dispersion (speed of temporal convergence in prices), the unit of analysis is the crop–market-group–day (crop–market–day).

By definition, each state will have its own crop–market-groups, with no overlap between states, and a crop can have multiple market-groups open on any day, each in a different state. We note that the same crop–market could be in multiple crop–market-groups because some markets are open on some days of the week and not others. As an example, consider the case where Market A is open Monday, Wednesday, and Friday and Markets B and C are open every weekday. On Monday, Wednesday, and Friday the market-group is "A-B-C". However, on Tuesday and Thursday, when Market A is closed, the market-group is "B-C'.' Since the market-group for a crop can differ from one day to the next, in contrast to most studies of geographic price dispersion, which have focused on situations where all markets under investigation were open in every time period (such as Jensen 2007, Aker 2008), our dataset is an unbalanced panel.

4.2. Variables

The main dependent variable is geographic price dispersion, which we measure using the coefficient of variation (CV) of prices. For robustness, we also use the volume-weighted coefficient of variation ($VWCV$) and the normalized range of prices (R), also referred to as the percentage price difference.

The CV_{ckt} is the ratio of the standard deviation of prices to the average price for crop c across markets in market-group k on day t. The CV is widely used to quantify geographic price dispersion (e.g. Sorensen 2000, Baye et al. 2006, Jensen 2007, Ghose and Yao 2011, Overby and Forman 2015). Unlike other measures that are sometimes used, such as the standard deviation, the CV is independent of the unit of measurement of prices, allowing meaningful comparison across crops that trade at different price levels or under different units of measurement (for example, volume vs. weight) and is independent of price inflation, which reached 12% in 2010 (Bartsch et al. 2010).

While the CV has many advantages, it weights all markets equally, irrespective of their traded volumes. To account for differences in trading volumes among markets we also estimate the $VWCV$, where the standard deviation is based on price deviations from a volume-weighted mean price. Intuitively, this measure is equivalent to computing the CV over all units sold. We also use a third measure of price dispersion: the difference between the highest price $P_{ckt}^{\max}$ and the lowest price $P_{ckt}^{\min}$, divided by the average price $\bar{P}_{ckt}$ in the crop–market-group–day, $R_{ckt} = (P_{ckt}^{\max} - P_{ckt}^{\min})/\bar{P}_{ckt}$. The (normalized) range considers only the most extreme prices in a crop–market-group on any day and been used extensively in the literature (e.g. Sorensen 2000, Ghose and Yao 2011). We normalize the range by dividing by the average price to have a measure that allows for comparisons across different crops and is independent of inflation.

Table 1 shows an example of the construction of the dependent variables, starting with observations at the crop–market–day level, which are converted to the crop–market-group–day unit of analysis. The table shows two separate but overlapping crop–market-groups for a single variety of

Table 1 Example of creating crop–market-group–day data from crop–market–day data

Raw crop–market–day data					Aggregate crop–market-group–day data						
Crop	Market	Date	Price	Volume	Crop	Market-group	Date	$\bar{P}$	CV	$VWCV$	R
Onion	Lasalgaon	Aug 30	458	12,500	Onion	A	Aug 30	455.5	0.151	0.057	0.351
Onion	Malegaon	Aug 30	360	600							
Onion	Mumbai	Aug 30	520	14,550							
Onion	Pimpalgaon	Aug 30	484	18,750							
Onion	Lasalgaon	Aug 31	481	12,000	Onion	B	Aug 31	493.67	0.046	0.033	0.061
Onion	Mumbai	Aug 31	510	6,500							
Onion	Pimpalgaon	Aug 31	480	13,500							

$\bar{P}$, CV, $VWCV$ and R represent the average price, coefficient of variation, volume-weighted coefficient of variation, and normalized range of prices of the crop–market-group–day, respectively.

onion in the state of Maharashtra for two days, August 30 and August 31, 2010. The raw crop–market–day price data is presented on the left and the aggregated crop–market-group–day data is on the right. In this example, c represents onion and k and t are indices representing market-group A and August 30, 2010 or market-group B and August 31, 2010, respectively.

4.3. Filters

For the main analysis we use price data for crops that have active RML subscribers from August 22, 2010 to November 8, 2010, for a total of 79 days and 235,309 individual crop–market–day observations. While a longer time window would have allowed a better estimation of any systematic time-invariant differences across crops and market-groups, we have chosen a relatively short time window to reduce the impact of time-varying factors such as seasonality and changes in infrastructure, regulations, or preferences.

Before converting the data from the crop–market–day level to the crop–market-group–day unit of analysis, we need to address some data issues. First, we find it necessary to exclude crop–markets that trade in a certain crop only occasionally. These crop–markets can generate crop–market-groups that appear in the data set infrequently. For the main analysis we remove all crop–markets with fewer than 10 crop–market–days in the study period. Since having 10 crop–market–day observations corresponds roughly to being open at least once a week, markets that trade less frequently are unlikely to be important markets for a particular crop. This eliminates 2,878 crop–market–days (1.22% of all crop–market–days). Second, we exclude data from any state i on date t that has only one crop–market open, as on that day it is not possible to calculate geographic price dispersion. This excludes 8,363 crop–market–days (3.55% of all crop–market–days).

We convert the remaining crop–market–day observations to the crop–market-group–day unit of analysis using the process described in the example in Table 1, resulting in 22,473 crop–market-group–day observations with between one and 62 observations in each crop–market-group. From these we remove 5,478 crop–market-groups that have only one observation, i.e., one day of data, as at least two observations are needed to estimate crop–market-group fixed effects. We remove

Table 2 Summary statistics and correlation matrix of crop–market-group–day data

	Overall		**Pre-ban**		**During-ban**		**Post-ban**		**Correlations**					
Variable	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	(1)	(2)	(3)	(4)	(5)	(6)
(1) Normalized Price	1.000	0.466	0.997	0.483	1.042	0.496	0.986	0.436	1	-0.267	-0.066	-0.011	-0.080	-0.075
(2) Normalized Volume	1.000	0.771	0.982	0.768	0.992	0.797	1.021	0.763	-0.267	1	-0.015	-0.084	-0.050	0.025
(3) CV	0.184	0.193	0.182	0.191	0.190	0.243	0.183	0.170	-0.066	-0.015	1	0.821	0.851	-0.014
(4) $VWCV$	0.137	0.191	0.135	0.180	0.144	0.241	0.136	0.178	-0.011	-0.084	0.821	1	0.776	-0.012
(5) R	0.474	0.775	0.473	0.762	0.506	1.212	0.463	0.522	-0.080	-0.050	0.851	0.776	1	0.028
(6) RML $subscribers$	221	1,071							-0.075	0.025	-0.014	-0.012	0.028	1
Observations	11,162		4,618		1,870		4,674							
Crops	159		159		159		159							
Crop–market-groups	601		601		601		601							
Days	75		32		12		31							

Normalized Price (Volume) refers to the price (volume) divided by the average price (volume) for that crop over all market-groups and days. This normalization is necessary for meaningful comparisons across crops as different crops trade in different units of price and/or volume. Note that by construction, the average Normalized Price (Volume) over the whole dataset is 1. CV, $VWCV$, and R stand for coefficient of variation, volume-weighted coefficient of variation, and normalized price range, respectively. RML $subscribers$ is the total number of subscribers that receive price information about markets in a crop–market-group, averaged over the time window of our study. By construction it does not change over time.

another 5,833 observations that come from crop–market-groups that did not have at least one observation in each stage of the ban (before, during, and after). We do so because it would be difficult to identify the impact of RML on price dispersion throughout the ban for crop–market-groups that do not have an observation at each ban stage. The final dataset contains 11,162 crop–market-group–day observations for 159 crops and 75 days. Summary statistics of the final dataset are presented in Table 2. In §6.4 we investigate whether the results are sensitive to the time window used for the analysis or any of these filters.

5. Actionability of RML Information

Before we proceed with the main analysis it is worth investigating whether the price information provided by RML is of any practical value to market participants. After all, by the time the information is collected, verified for quality assurance purposes and transmitted via text message to subscribers, it is often not actionable until the next day. Therefore, it is not clear if subscribers can benefit from having access to this information, which, by the time they can actually visit the market, is going to be a day old. Furthermore, the information may not be actionable if subscribers, who in the vast majority of cases are rural farmers, have to travel a prohibitively long distance in order to take advantage of the information they receive through RML.

Since we do not know which markets RML subscribers chose to visit or they would have visited had they not had access to RML, it is difficult to measure the direct value of the information. However, we can perform the following check. For every subscriber and crop we use geolocation data to determine the price at which the subscriber would be able to sell at his closest market on each day in the period from August 22, 2010 to November 8, 2010. We then compare this price to the price at the market that had the highest price the previous day among the markets he had subscribed to for that crop. We report the difference in price between these two markets in Row (1) of Table 3. This suggests that if instead of selling at his closest market, a subscriber sold his crop

Table 3 Actionability of RML price information

Variable	Closest Market		Average of Two Closest Markets	
	Price Diff. (%)	Distance Diff. (Miles)	Price Diff. (%)	Distance Diff. (Miles)
(1) Subscribed Max (t-1)	18.7	75.1	16.4	65.0
(2) Subscribed Max within 100 miles (t-1)	6.5	20.4	5.3	11.4
(3) Subscribed Max within D_c miles (t-1)	7.9	19.8	6.3	10.5
(4) Subscribed Max (t)	20.0	75.1	17.6	65.0
(5) Subscribed Max within 100 miles (t)	7.4	20.4	6.2	11.4
(6) Subscribed Max within D_c miles (t)	8.6	19.8	7.0	10.5

All numbers are significantly different from zero. D_c denotes the average market-to-subscriber distance for crop c. The table reports price difference and additional distance traveled between market Y and market X. In the first two columns, market Y is closest to the subscriber. In the last two columns, Y is the average of the two markets closest to the subscriber. Market X is one of the markets for which they receive RML information. In Row (1) X is the market with the highest previous day price; (2) X is the market with the highest previous day price that is within 100 miles of the subscriber; (3) X is the market with the highest previous day price that is within D_c miles of the subscriber (4) X is the market with the highest price on the day, (5) X is the market with the highest price on the day that is within 100 miles of the subscriber, (6) X is the market with the highest price on the day that is within D_c miles. We use data from the period August 22, 2010 to November 8, 2010. The dataset contains a 62,503 subscribers in 1,165 unique locations for 156 crops and 969 markets (4,362 crop–markets).

at the market with the highest previous-day price among the markets in his RML subscription, he would receive, on average, 18.7% more. To do so, he would have to travel, on average, an additional 75.1 miles. If we restrict the maximum distance a farmer is willing to travel to 100 miles, then the price increase obtained by using the RML information would be, on average, 6.5% and the average additional distance traveled, 20.4 miles (see Table 3, Row 2). By comparison, if instead of having access to one-day-old information a farmer had access to instantaneous price information, the benefit would increase from 18.7% to 20.0% (or from 6.5% to 7.4% if travel were restricted to 100 miles) – see Table 3, Row 4 (5). The benefit is positive, on average, for 84% of subscribers. Results are qualitatively unchanged if we restrict the maximum distance traveled to the average market-to-subscriber distance for each crop, or if instead of making comparisons with the price at the closest market, we make comparisons with the average price of the two closest markets (see Table 3).

This analysis suggests that the RML-provided information is relevant to its subscribers. First, despite being one day old it still allows subscribers to capture 94%(=18.7%/20.0%) of the value associated with shifting supply (or demand) to exploit geographic price variation. This finding implies that prices of agricultural goods in India are, at least to some extent, serially correlated and this temporal correlation may be exploited by RML subscribers. We further investigate this in §7. Second, the distances subscribers need to travel are not prohibitive. The fact that this information is of practical value to subscribers, i.e., it allows them to better match supply with demand, is a necessary but not sufficient condition to argue that RML-provided information has an impact on geographic price dispersion. We investigate this in §6.

6. The Impact of RML Information on Geographic Price Dispersion

In this section we investigate whether price information, such as that provided by RML, has an effect on geographic price dispersion. To do so we exploit the text-message ban, which was an exogenous intervention completely unanticipated by RML management and subscribers. During the ban RML subscribers lost one source of information regarding the price of agricultural crops in nearby markets. Nevertheless, they still had access to all other means of information retrieval, including mobile phone ICT. One possibility is that the absence of this information had no effect on subscribers' decision-making process: subscribers were still able to choose the most advantageous markets to transact in because they were able to acquire the relevant information through alternative sources that were not disrupted. In this case one would expect that geographic price dispersion during the ban would be statistically similar to that before and after the ban. Another possibility is that the loss of information affected subscribers' ability to decide which market to visit; without RML subscribers made market-choice decisions based on criteria other than price (for example, distance) or based on incomplete or inaccurate price information they were able to acquire through other means. In this case, and to the extent that RML subscribers command significant enough quantities of goods to have a material impact on prices, geographic price dispersion should be higher during the ban compared to before or after it. A third possibility is that RML subscribers stopped trading altogether in at least a subset of markets during the ban, either due to fear that there would be riots (which is why the ban was imposed in the first place) or simply in anticipation of the information resuming at the end of the ban. If this is the case, geographic price dispersion may be different during the ban but, more importantly, one would expect to see systematic changes in the volumes transacted and in the concentration of volumes across different markets. We investigate these possibilities below.

6.1. The Baseline Model

We first examine the simplest possible model that allows to test whether price dispersion differs before, during, and after the bulk text message ban,

$$Y_{ckt} = \alpha_{ck} + \beta_1 DuringBan_t + \beta_2 PostBan_t + \epsilon_{ckt}, \tag{1}$$

where Y_{ckt} denotes the measure of geographic price dispersion (CV, $VWCV$ or R) for crop c in market-group k on day t, ϵ_{ckt} is the error term, α_{ck} are crop–market-group fixed effects, $DuringBan_t$ is a dummy variable that takes the value 1 if the date t falls within the text message ban (i.e., September 22 to October 4, 2010 inclusive), and $PostBan_t$ is a dummy variable that takes the value 1 if the date t falls after the ban was lifted (i.e., after October 4, 2010). Therefore, the coefficients β_1 and β_2 will measure how much higher, on average, geographic price dispersion is during and after the ban, respectively, relative to the pre-ban period.

The crop–market-group fixed effects α_{ck} in Model (1), which subsume crop, market-group, crop–market, and state fixed effects, play an important role in our identification strategy. More specifically they control for systematic differences in price dispersion across crops due to differences in perishability, usage, demand elasticity, population preferences, etc. They also control for systematic differences in price dispersion across market-groups due to differences in their spatial structure, population, incomes, habits, presence of terminal markets, road infrastructure, storage facilities, cost of transport, etc. Finally, they control for the interaction of crop and market-group effects; for example, a highly perishable crop may have a relatively high price dispersion on average but a lower price dispersion in a crop–market-group with good cold storage facilities. Due to the short time window of our study, we do not expect any of these factors to change in a materially meaningful way. In this simple model, we cannot include date fixed effects to control for systematic day-to-day variation in price dispersion, as they would be collinear with the variable of interest, $DuringBan_t$. We address this problem later.

Before we present the results of Model (1), we need to discuss the assumptions behind the variance–covariance matrix of the error term ϵ_{ckt}. One possibility would be to assume that errors are independent and identically distributed (iid). However, there is evidence, based on the residuals of (1), to suggest that this is not the case. A Wald test rejects the hypothesis that errors are homoskedastic across crop–market-groups ($\chi^2(601) = 1.5 \times 10^8; p < 0.001$) and an F-test rejects the hypothesis of no first-order autocorrelation ($F(1, 237) = 5.393; p = 0.021$). Therefore, the iid assumption would result in underestimating the standard errors of the estimated coefficients, leading to erroneous inference. To overcome this problem we cluster errors at two levels: the crop level and the date level (Cameron et al. 2011).[5] Clustering at the crop level allows for heteroskedastic errors across crops, as well as arbitrary correlation of errors across crop–market-groups and across dates within a crop. Note that clustering instead at the panel unit – the crop–market-group level, which is a lower level of aggregation – would not permit correlation in errors associated with different market-groups trading the same crop c. Allowing for this type of correlation is particularly important as there is some overlap in the markets across market-groups (e.g., in Table 1 the Lasalgaon market belongs to both market-groups A and B). Clustering on the date level allows for heteroskedastic errors across days, which is important as the ban may have impacted the error variance, and for errors associated with different crop–market-groups to be correlated on the same date. Moreover, the number of crops (159) and the number of days (75) each exceed the threshold of 50 suggested by Wooldridge (2006).

The results from Model (1) with CV as the measure of geographic price dispersion are presented in Column 1 of Table 4. We find that on average (over all crops and all market-groups) the CV during the ban is higher by 0.010 than before the ban, (p-value=12.8%) and by 0.009 (=0.010-0.001) than after the ban (p-value=21.4%). Also, the CV post-ban is not significantly different

[5] The two-way error clustering at the crop level and at the date level is quite different from one-way error clustering at the crop-date level; the latter imposes a much more restrictive variance–covariance matrix.

Table 4 **Model (1) regression results**

VARIABLES	Fixed Coefficients (1) CV	(2) $VWCV$	(3) R	(4) $\log(Volume)$	(5) $CV(Volume)$	Random Coefficients (6) CV Mean Coef.	St. Dev.
DuringBan	0.010	0.011*	0.036	0.021	0.010	0.012*	0.055
	(0.006)	(0.006)	(0.028)	(0.027)	(0.010)	(0.006)	(0.014)
PostBan	0.001	0.004	-0.005	0.042	-0.006)	0.004	0.045
	(0.006)	(0.005)	(0.013)	(0.029)	(0.011)	(0.005)	(0.010)
Observations	11,162	11,162	11,162	11,162	11,162	11,162	
Adjusted R-Squared	0.640	0.686	0.495	0.945	0.843		
Pseudo loglikelihood						8709.97	
No. of crop–market-groups	601	601	601	601	601	601	

*** p<0.01, ** p<0.05, * p<0.1, crop–market-group–day panel regressions with crop–market-group fixed effects. Standard errors, with two-way clustering at the crop level and at the date level in Columns (1)-(5) and with one-way clustering at the crop level in Column (6), reported in parentheses. CV, $VWCV$, and R represent the coefficient of variation, volume-weighted coefficient of variation, and normalized range of prices, respectively. $Volume$ and $CV(Volume)$ denote the average volume and the coefficient of variation of volume in a crop–market-group, respectively. The panel includes crop–markets with at least 10 observations and crop–market-groups composed of at least two markets and with at least one observation before, during, and after the ban.

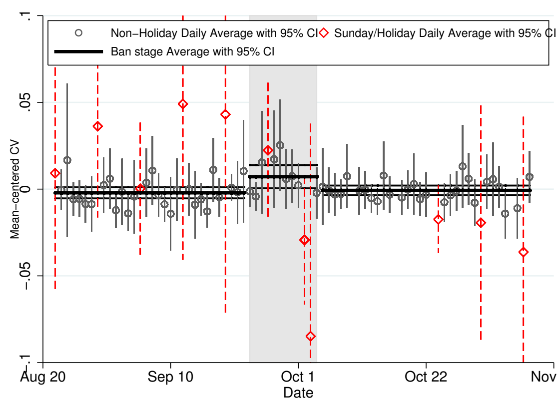


Figure 1 **Average CV for each day (ban stage) together with 95% C.I., after subtracting crop–market-group mean CV. Diamond dots denote Sundays or public holidays.**

from its pre-ban level. This relationship is shown graphically in Figure 1. Each dot represents the average price dispersion (CV) on a given day after subtracting out crop–market-group means, along with a 95% confidence interval. A positive number denotes daily price dispersion that is, on average, higher than that for the crop–market-group mean. The horizontal lines show the average pre-, during- and post-ban price dispersion after subtracting out the crop–market-group means, along with their 95% confidence intervals. As can be seen in the figure, there are no significant time trends before, during or after the ban. Furthermore, the daily average price dispersion before and after the ban is roughly equally likely to be higher or lower than the crop–market-group mean (it is higher 23 out of 63 days); during the ban it is more likely to be higher (nine out of 12 days). The results of Model (1) for $VWCV$ and R are shown in Columns 2 and 3 of Table 4. Although the differences between the during- and pre-/post-ban periods are not significant at conventional levels for most of these measures, if we consider that the reported effect is an average over all crops and all markets covered by RML, including those with few subscribers, the sign and relative magnitude of the coefficient are encouraging.

6.2. Difference-in-differences model

Naturally, if the RML-provided information has an effect on price dispersion, one would expect the impact of the ban to be greater in crop–market-groups with relatively higher exposure to RML than in those where RML does not have a substantial presence. In this section we extend the Model of (1) to measure this differential impact.

To determine the exposure of different crop–market-groups to RML information we need to construct appropriate penetration measures. Our measures use information on the number of subscribers in each crop–market-group. This number is known with a high degree of accuracy from the data provided by RML. Since there is little variation in RML subscriptions over the short time window centered around the time of the ban, we exploit variation in the number of subscribers across crop–market-groups. To construct a sensible penetration measure, besides the number of subscribers in each crop–market-group, we need to know the potential market size of the crop–market-group (e.g., how many farmers transact in total) as well as how influential RML subscribers are vis-à-vis other market participants who do not have access to RML information (e.g., how much volume RML subscribers transact relative to non-subscribers). These last two figures are difficult to estimate.[6] Therefore, any measure constructed based on the number of subscribers will capture RML penetration with error. Although this error is unlikely to be correlated with the ban (which is exogenous), it may still lead to a bias in the coefficients of interest, i.e., cause measurement-error bias (see, for example Tambe and Hitt 2013). In an attempt to somewhat reduce the concerns that this naturally raises we define three complementary measures of penetration. Although they are not independent from one another, they make different adjustments that capture different aspects of penetration. We show that our results are qualitatively the same, with all measures. In §6.4 we discuss the limitations of this approach further.

The first measure of penetration we use is the number of RML subscribers in a crop–market-group divided by the total volume transacted at the crop–market-group, which we call subscribers per

[6] We were unable to accurately estimate these two figures due to data availability constraints. More specifically, we tried to estimate the total number of farmers in a crop–market-group using data reported in the latest (2006) agricultural census. However, the census collected data using different crop classifications than those used by RML and at a much coarser level of geographic aggregation than the market-group level. We also tried to estimate the volume transacted by RML subscribers using (i) data collected by RML when new subscribers activated their subscriptions (this self-reported data included information on acreage under cultivation); however, over 95% of subscribers did not report their land size, and (ii) data from Farmer Interest Group surveys; however these surveys were conducted at a coarser level than the crop–market-group.

volume (SV). The assumption behind this measure is that the volume transacted is a good proxy for the potential size of the market (i.e., the total number of market participants). By dividing the number of subscribers by the volume transacted at the crop–market-group level, we are essentially assuming that if two market-groups have the same number of RML subscribers for a particular crop, RML is relatively more influential (on average) in the market-group where the transacted volume is lower. With this measure we can rank market-groups within a crop in terms of RML influence, albeit with error.

Since the volumes of different crops are not comparable (e.g., some crops trade in units of volume, others in units of weight), a drawback of this measure is that it cannot be used to compare market-groups across crops. In an attempt to overcome this shortcoming we define two additional penetration measures – normalized subscribers-per-volume (NSV) and subscribers-per-Rupee (SR) – that allow across-crop comparisons.

The second measure, NSV, is simply a normalization of the SV measure attained by dividing SV by the average number of subscribers per volume for each crop. Crop–market-groups that rank high on this measure (i.e., those in its highest quartile) are those that have a high number of subscribers-per-volume relative to the average number of subscribers-per-volume for the crop in question and belong to crops that have relatively wide variation in the number of subscribers-per-volume. In contrast to the SV measure, this measure has the drawback that crop–market-groups that belong to crops that have consistently high (or low) penetration across all market-groups will be classified as medium-penetration.

For the third measure of penetration, we start from the observation that there is systematic variation in price across crops and we find evidence that the number of farmers per volume transacted is larger for crops that trade at a higher price.[7] This suggests that crops trading at a higher average price are associated with a larger number of farmers per volume transacted compared to crops with a lower average price. Therefore, the same number of RML subscribers per volume would translate into smaller penetration for crops with higher price. (A possible explanation for this observation is that more farmers find it sufficiently profitable to grow and trade small quantities of crops that trade at a higher price than for those trading at a lower price. We should emphasize that this is only true on average, i.e., there might be some high-priced crops for which the scale economies

[7] The evidence comes from combining data from the latest (2006) agricultural census, which reports the number of farmers for 397 crop–states, with RML data from the whole year of 2010, which allows us to calculate the average daily volume transacted and the (volume-weighted) average price of each crop–state. (These averages are taken over all crop–market-groups within the state.) Assuming that the number of farmers is relatively stable from 2006 to 2010, we then regress $\log(NumberOfFarmers/Volume)$ on $\log(AveragePrice)$ with state fixed effects. The results support the assumption that higher prices are associated with more farmers per unit of volume transacted. The coefficient of $\log(AveragePrice)$ is positive (0.922) and significant (p-value$<$0.1%; adjusted R-squared=0.290). A further regression, where instead of estimating the number of farmers using the agricultural census data we use an estimate based on Farmer Interest Group data, produced similar results.

require farmers to produce in high volumes to be competitive.) The third measure of penetration we use, subscribers per Rupee (SR), is defined as number of RML subscribers in a crop–market-group divided by the product of the total transaction volume in that crop–market-group and the volume-weighted average price of the crop (measured in Rupees) across all crop–market-groups. As noted above, the main advantage of this measure over the SV measure is that it allows for comparisons across crops.

Details of the exact calculations of all three measures can be found in the Appendix. We note that the three measures have a high degree of overlap but they are not identical. The Spearman rank correlation coefficient between SR–NSV, NSV–SV, and SV–SR is 60%, 77%, and 52%, respectively. Furthermore, by construction, the rankings of crop–market-groups within a crop using SR, NSV, and SV are all identical. For the main analysis we use the four quartiles of these measures. We refer to crop–market-groups in the top (bottom) quartile as high- (low-) penetration markets.[8] As in the case of the SV variable, we use the quartiles of these two measures.

We proceed by fitting a difference-in-differences model in which we interact the quartiles of the three penetration measures with the $DuringBan_t$ and $PostBan_t$ variables. Specifically, we estimate the econometric model:

$$Y_{ckt} = \alpha_{ck} + \delta_t + \sum_{i=2}^{4} \theta_i DuringBan_t \times Quartile\text{-}i_{ck} + \sum_{i=2}^{4} \zeta_i PostBan_t \times Quartile\text{-}i_{ck} + \epsilon_{ckt}, \qquad (2)$$

where Y_{ckt}, ϵ_{ckt}, $DuringBan_t$, $PostBan_t$ and α_{ck} are as defined in Model (1). In this specification the interaction between the $DuringBan_t$ (or $PostBan_t$) variable and the lowest quartile of the penetration variable ($Quartile\text{-}1_{ck}$) is the omitted category. Therefore, the coefficients of the other three quartiles can be interpreted in the usual difference-in-differences manner. For example, the coefficient of the interaction term $DuringBan \times Quartile\text{-}4_{ck}$ is the difference between the changes in price dispersion during the ban for the highest penetration quartile relative to before the ban and changes in price dispersion during the ban for the lowest penetration quartile relative to before the ban, after subtracting out fixed effects.

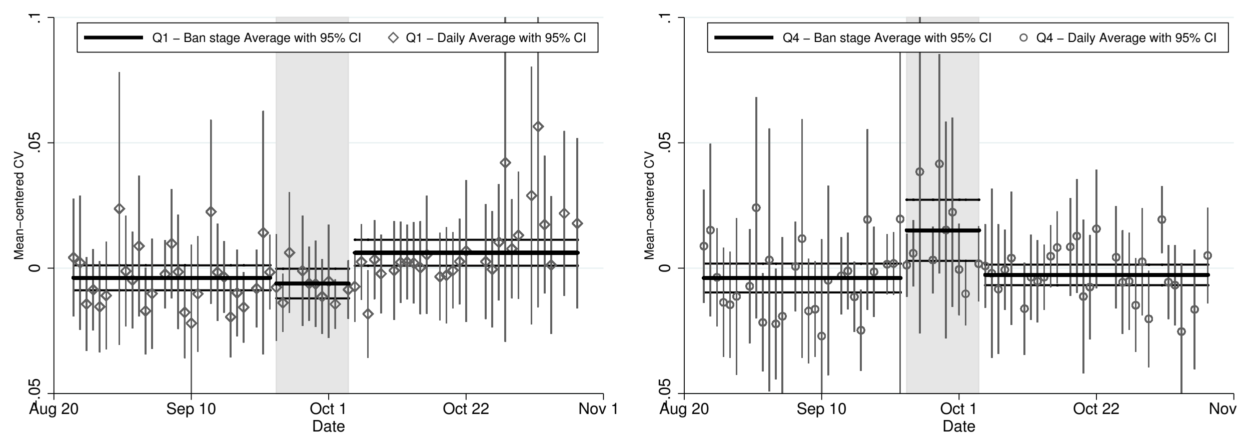
In this specification, in contrast to that of Model (1), we are interested in the differential variation of the impact of the ban in markets where RML has a high level of penetration relative to those where it has low penetration, measured by the coefficient of the interaction term $DuringBan_t \times Quartile\text{-}4_{ck}$, rather than the aggregate effect of the ban measured by the coefficient of the $DuringBan_t$ dummy variable. Therefore, in this specification it is possible to replace the $DuringBan_t$ and $PostBan_t$ dummy variables with a set of more detailed date fixed effects

[8] In constructing the quartiles we follow the convention of placing all ties in the same quartile starting from the quartile of highest penetration. This procedure leads to quartiles that have an unequal number of crop–market-groups. Results using other conventions for breaking ties yield similar results.

Table 5 **Model (2) regression results**

VARIABLES	(1) CV	(2) $VWCV$	(3) R	(4) CV	(5) $VWCV$	(6) R	(7) CV	(8) $VWCV$	(9) R
$DuringBan \times Quartile$-2	0.004	0.007	0.017	-0.009	-0.004	-0.018	0.001	-0.000	0.006
	(0.011)	(0.010)	(0.024)	(0.011)	(0.009)	(0.021)	(0.010)	(0.009)	(0.034)
$DuringBan \times Quartile$-3	0.023	0.027	0.161	0.001	-0.002	0.010	0.032	0.025	0.168
	(0.022)	(0.019)	(0.131)	(0.012)	(0.010)	(0.026)	(0.025)	(0.021)	(0.114)
$DuringBan \times Quartile$-4	0.022**	0.023**	0.073***	0.034*	0.028	0.160	0.047**	0.034	0.205*
	(0.011)	(0.010)	(0.021)	(0.018)	(0.018)	(0.098)	(0.020)	(0.021)	(0.107)
$PostBan \times Quartile$-2	-0.006	0.001	-0.011	-0.012	-0.006	-0.006	-0.013	-0.004	0.019
	(0.015)	(0.010)	(0.036)	(0.013)	(0.011)	(0.027)	(0.019)	(0.014)	(0.053)
$PostBan \times Quartile$-3	-0.026	-0.005	-0.019	0.005	0.004	0.038	-0.028	-0.012	-0.060
	(0.016)	(0.011)	(0.023)	(0.017)	(0.015)	(0.038)	(0.024)	(0.016)	(0.063)
$PostBan \times Quartile$-4	-0.009	0.009	-0.027	-0.009	-0.007	-0.022	-0.004	0.003	0.033
	(0.012)	(0.011)	(0.032)	(0.011)	(0.012)	(0.029)	(0.012)	(0.011)	(0.029)
Observations	11,162	11,162	11,162	11,162	11,162	11,162	6,911	6,911	6,911
Adjusted R-Squared	0.640	0.686	0.494	0.641	0.686	0.495	0.614	0.582	0.461
No. of crop–market-groups	601	601	601	601	601	601	402	402	402

*** p<0.01, ** p<0.05, * p<0.1, crop–market-group–day panel regressions with date and crop–market-group fixed effects. Standard errors, with two-way clustering at the crop level and at the date level are reported in parentheses. CV, $VWCV$, and R represent the coefficient of variation, volume-weighted coefficient of variation, and normalized range of prices, respectively. In the construction of the panel we include crop–markets with at least 10 observations and crop–market-groups composed of at least two markets and with at least one observation before, during, and after the ban. All columns include crop–market-group fixed effects and date fixed effects. In columns 1–3, 4–6 and 7–9 crop–market-groups are ranked in quartiles according to the NSV, SR and SV variables, respectively. In the models of Columns 7–9 we exclude crops with three or fewer market-groups.



(a) $Quartile$-1 (b) $Quartile$-4

Figure 2 **Average CV for each day (ban stage) together with 95% C.I., after subtracting crop–market-group mean CV.**

δ_t. Essentially, these date fixed effects serve as non-parametric controls for the temporal variation in price dispersion that is common across all crop–market-groups. Also, note that in this model we cannot estimate a coefficient for the quartiles of the penetration variable directly as they are collinear with, and completely subsumed by, the crop–market-group fixed effects.

We present the results of Model (2), with CV as the measure of geographic price dispersion and penetration measured using the NSV variable, in Column 1 of Table 5. The coefficient of the interaction between the fourth quartile and the during-ban variable ($DuringBan_t \times Quartile\text{-}4_{ck}$) is positive and significant (coefficient=0.022, p-value=4.2%). This suggests that price dispersion was indeed higher during the text message ban in those crop–market-groups that belong to the highest quartile of RML penetration vis-à-vis those crop–market-groups that belong to the lowest quartile of penetration. When the ban is lifted, the difference between the CV of the highest-quartile crop–market-groups and the lowest-quartile crop–market-groups is no different to what it was before the ban. This is indicated by the coefficient of the interaction between the fourth quartile and the post-ban variable ($PostBan_t \times Quartile\text{-}4_{ck}$) of -0.009 (p-value=43.3%). Further, the difference in price dispersion between the first and the fourth quartile during the ban is higher than post ban (difference 0.032, p-value=0.8%). These relationships are illustrated in Figure 2, which contrasts the price dispersion of the first and fourth quartile of the penetration variable. We note that after subtracting the crop–market-group means, the average price dispersion of the fourth quartile (right figure) is not significantly different from the average price dispersion of the first quartile (left figure) before or after the ban, but it is during the ban (the two confidence intervals of the price dispersion during the ban do not overlap). Comparing price dispersion on individual days, crop–market-groups that belong to the fourth quartile of penetration are roughly equally likely to have a higher or lower price dispersion than those in the first quartile before or after the ban (they are higher on 24 out of 55 days); during the ban they are higher every day (nine out of nine days). Returning to Column 1 of Table 5, the differential impact of the ban on price dispersion in the third quartile of penetration is also positive and similar in magnitude to that in the fourth quartile, but is not statistically significant (coefficient=0.023, p-value=29.9%), while the impact of the ban on the second quartile does not differ from the impact of the ban on the lowest quartile (coefficient=0.04, p-value=68.5%).

Columns 2 and 3 of Table 5 show the results of Model (2), with $VWCV$ and R as the measures of geographic price dispersion and where penetration is still measured with the NSV variable. Columns 4–6 and 7–9 show the results using SR and SV to measure penetration. Note that the SV measure ranks market-groups according to number of subscribers-per-volume within a crop. To avoid dividing crops with three or fewer market-groups into quartiles, we exclude from the analysis 4,251 observations that come from such crops. The results are qualitatively similar.

Importantly, as can be seen from all models in Table 5, significant price dispersion increases occurred only during the ban and only in those crop–market-groups where RML had the highest penetration. By contrast, crop–market-groups where RML had a limited presence exhibited no significant difference in price dispersion during the ban relative to the pre-ban or post-ban periods.

To get a better sense of the magnitude of the impact that RML has on price dispersion, we note that the 0.022 average increase in price dispersion observed during the ban in those crop–market-groups where RML has the highest penetration is approximately 12.0% of the average CV (which was 0.184 – note that this is the average dispersion over the entire dataset including the period during the ban, making this a conservative estimate). Similar calculations that use $VWCV$ and R as the measures of price dispersion suggest that the impact of RML in reducing price dispersion is, on average, 16.8% and 15.4% for $VWCV$ and R, respectively, using the NSV measure. Coefficients using SV and SR are larger and would therefore lead to larger estimates of the effect of RML on price dispersion.

Further to the observation that in the absence of RML information the average price dispersion increases, we investigate how the absence of RML information affects the distribution of price dispersion. Using the Wilcoxon–Mann–Whitney test for first-order stochastic dominance we find that deviations from the average CV during the ban do not stochastically dominate those before the ban for crop–market-groups in the first quartile (p-value=40.5%) but do for crop–market-groups in the fourth quartile (p-value=1.3%) using NSV to measure penetration. Results with price dispersion measured by $VWCV$ and R are similar.

Since all other methods of obtaining price information were unaffected by the text message ban, this increase in price dispersion can be attributed to the loss of RML price information alone. Therefore, we can conclude that the information RML provides reduces geographic price dispersion by 12.0–16.8% more on average in those markets in the top quartile compared to those in the bottom quartile of penetration. Previous research has shown that the introduction of mobile phone infrastructure can reduce geographic price dispersion by 10–20% on average (Aker 2008, 2010). Although not directly comparable because of the different contexts and applications, our research suggests that access to timely and accurate information, such as that provided by RML, can reduce geographic price dispersion by an additional amount that is of similar magnitude to the reduction resulting from access to mobile phone technology.

In the next two sections we investigate alternative hypotheses that might explain the main finding, together with a number of robustness checks. Unless otherwise stated, we conduct the investigation in the context of the difference-in-differences specification of Model (2) with CV as the measure of geographic price dispersion and NSV as the measure of RML penetration. Results obtained using the other two measures of price dispersion and penetration are qualitatively similar.

6.3. Alternative Explanations

Our identification strategy relies on the differential impact of the ban on price dispersion in crop–market-groups where RML has a relatively high vs. low level of penetration. This difference-in-differences methodology should alleviate concerns that the increase in price dispersion during the

Table 6 Tests for alternative explanations

	Fixed Coefficients							Random Coefficients	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
VARIABLES	log($Volume$)	$CV(Volume)$	log(P)	CV	CV	CV	CV	CV	
								Mean Coef.	St. Dev.
$DuringBan \times Quartile$-2	0.003	0.013	0.021	0.001	0.029**			0.005	0.029
	(0.055)	(0.032)	(0.020)	(0.017)	(0.015)			(0.011)	(0.014)
$DuringBan \times Quartile$-3	-0.049	-0.028	0.014	0.012	0.031*			0.018	0.116
	(0.061)	(0.027)	(0.024)	(0.018)	(0.017)			(0.017)	(0.041)
$DuringBan \times Quartile$-4	0.067	-0.017	0.018	-0.001	0.033***	0.015	0.021**	0.022*	0.104
	(0.056)	(0.029)	(0.018)	(0.016)	(0.012)	(0.010)	(0.011)	(0.012)	(0.038)
$PostBan \times Quartile$-2	0.023	0.026	0.012	-0.021	-0.000			-0.004	0.053
	(0.068)	(0.035)	(0.035)	(0.021)	(0.013)			(0.013)	(0.010)
$PostBan \times Quartile$-3	-0.016	0.026	0.062	-0.019	-0.008			-0.012	0.084
	(0.085)	(0.035)	(0.041)	(0.016)	(0.014)			(0.015)	(0.089)
$PostBan \times Quartile$-4	0.061	0.012	0.059	-0.032	0.004	-0.001	-0.010	-0.012	0.054
	(0.076)	(0.037)	(0.037)	(0.020)	(0.013)	(0.010)	(0.012)	(0.012)	(0.011)
Observations	11,162	11,162	11,162	5,391	9,050	11,162	6,583	11,162	
Adjusted R-Squared	0.946	0.842	0.989	0.762	0.779	0.640	0.570		
Pseudo loglikelihood								8912.47	
Start Date	22-Aug-10	22-Aug-10	22-Aug-10	22-Aug-09	22-Aug-10	22-Aug-10	22-Aug-10	22-Aug-10	
End Date	8-Nov-10	8-Nov-10	8-Nov-10	8-Nov-09	8-Nov-10	8-Nov-10	8-Nov-10	8-Nov-10	
No. of crop–market-groups	601	601	601	285	502	601	359	601	

*** p<0.01, ** p<0.05, * p<0.1, crop–market-group–day panel regressions with date and crop–market-group fixed effects. Standard errors, with two-way clustering at the crop level and at the date level in Columns (1)-(7) and with one-way clustering at the crop level in Column (8), reported in parentheses. CV, $Volume$ and P denote the coefficient of variation of prices, the average volume and average price in a crop–market-group, respectively. $CV(Volume)$ is the coefficient of variation of volume in a crop–market-group. In the construction of the panel we include crop–markets with at least 10 observations and crop–market-groups composed of at least two markets and with at least one observation before, during, and after the ban. Column 4 estimates Model (2) using data from 2009 instead of 2010. Column 5 uses density-based instead of state-based clustering. Column 6 estimates Model (2) using all of the data and merging the three lowest quartiles. Column (7) estimates Model (2) excluding data from the second and third quartiles.

ban is not causally linked to RML. For example, concerns that the quality of the data-collection process may have deteriorated during the ban (e.g, because RML market reporters, knowing that the information would never reach customers, were less diligent than usual) cannot explain the differential impact of the ban on the high- vs. low-penetration crop–market-groups. If data collection errors crept in during the ban, this would have affected high- and low-penetration crop–market-groups equally. Nevertheless, there are two potential alternative explanations that we believe are worth exploring further: (a) a change in liquidity due to a change in the trading behavior of market participants during the ban and (b) extreme weather phenomena during the ban.

First, it may be that during the ban market participants stopped trading. This may have happened as a direct result of the civil unrest that was expected to break out at the time of the text message ban. (Note that for civil unrest to be compatible with the results it also needs to be the case that market participants felt more unsafe visiting the market-groups in the top quartile of RML penetration compared to those in the bottom quartile.) Alternatively, this may have happened because market participants, especially those subscribing to RML, decided to postpone trading until after the ban was lifted. In either case, the change in price dispersion observed during the ban would be due to a change in liquidity rather than an increase in informational frictions.

We do not believe this is the case as, fortunately, no civil unrest occurred before or after the Ayodhya verdict (see §3.3). Furthermore, if the effect was due to market participants not trading, we would also expect to observe systematic changes in the pattern of transactions during the ban. In particular, we would expect to see (a) a reduction in the average volume transacted during the ban and/or (b) a systematic shift in the volume transacted from some, presumably unsafe, markets within a market-group to other, presumably safer, markets within the market-group. We check whether (a) is the case by estimating Models (1) and (2) with $\log(Volume)$ as the dependent variable. The results presented in Column 4 of Table 4 show that volumes were not significantly different during the ban compared to before or after the ban. Similarly, the results presented in Column 1 of Table 6, show that volumes were not significantly different in high-penetration areas relative to low-penetration areas during the ban. We check if (b) is the case by estimating Models (1) and (2) with the coefficient of variation of volume as the dependent variable. We do not see any significant difference in the CV of volume within a crop–market-group (Column 5 of Table 4 and Column 2 of Table 6). Together, these constitute evidence against this alternative explanation.

A second alternative explanation is that an extreme weather event may have occurred in areas of high RML penetration during the ban, preventing market participants from transacting at these markets. However, if this was the case, we would expect volumes to have changed as discussed in the previous paragraph. The fact that volumes transacted did not change in any systematic way suggests that this is unlikely to be the case.

It is also worth examining the impact of the ban on average price levels as opposed to price dispersion. We use the logarithm of price to account for the fact that different crops trade at different price levels/volume units. Column 3 of Table 6 shows that differences in prices during the ban were not statistically different from zero for any quartile of RML penetration.

We note that Columns 1 and 3 of Table 6 show that average volumes and prices in a crop–market-group were not significantly different during the ban. As a further check, not reported here, we verified that volumes and prices measured at the crop–market level, rather than the average of the crop–market-group, were also unchanged.

6.4. Additional Robustness Checks

In this section we perform several robustness checks. First, one concern with Models (1) and (2) is that they assume that the impact of the RML information ban on price dispersion is homogenous across all crops (and all crop–market-groups). If this assumption does not hold, the average effect of the information provided by RML on price dispersion during the ban may actually be different than that estimated using Models (1) and (2): a phenomenon known as heterogeneity bias (Hsiao 2003). Econometrically, an efficient and consistent way to account for this heterogeneity is to estimate a random coefficients specification that explicitly models the variation in treatment across crops (see Hsiao (2003) for more information). Under the random coefficients specification, Model (1) becomes:

$$Y_{ckt} = \alpha_{ck} + \beta_{1c} DuringBan_t + \beta_{2c} PostBan_t + \epsilon_{ckt},$$

where $\beta_{ic} = \beta_i + \eta_{ic}$, and Model (2) becomes:

$$Y_{ckt} = \alpha_{ck} + \delta_t + \sum_{i=2}^{4} \theta_{ic} DuringBan_t \times Quartile\text{-}i_{ck} + \sum_{i=2}^{4} \zeta_{ic} PostBan_t \times Quartile\text{-}i_{ck} + \epsilon_{ckt},$$

where $\theta_{ic} = \theta_i + \nu_{ic}$, $\zeta_{ic} = \zeta_i + \xi_{ic}$, ϵ_{ckt} is the error term, and η_{ic}, ν_{ic}, and ξ_{ic} are random variables that capture crop-specific heterogeneity. We are interested in estimating β_i, θ_i, and ζ_i while specifically controlling for these crop-specific deviations from the coefficient of interest. As is common in the application of random coefficient models (Hsiao 2003), all random coefficients are assumed to have a time-invariant variance–covariance matrix and are assumed to be independent of each other. In this specification, we can allow for clustered errors at the crop level only, as opposed to the two-way clustered errors used in the main analysis. Column 6 of Table 4 and column 8 of Table 6 show the results of estimating the crop-specific random coefficients models: the left column shows the average effect while the right column shows the standard deviation of the crop-specific random effects η_{ic}, ν_{ic}, and ξ_{ic}. We note that a likelihood ratio test confirms that the random-coefficients model fits the data better than a fixed-coefficients model (in both models $p < 0.001$). The main conclusion, however, is that the results are robust to allowing heterogeneous treatment. The average impact of the ban across all crops is positive and significant for those crop–market-groups in the highest quartile of RML penetration. The results of a similar model that allows for heterogeneity at the crop–market-group level are qualitatively similar (not shown here).

Second, we check that there is nothing special about the specific time period (September 23–October 4) in which the ban occurred by repeating our analysis using data from 2009 instead of 2010. Column 4 of Table 6 shows, reassuringly, that there was no significant change in price dispersion during the same time period in 2009.

Third, up to this point we have examined price dispersion in "groups" of markets that were located within state boundaries. This specification, although justified by linguistic, cultural, and regulatory frictions in cross-state trading, has the drawback of precluding markets that are in close physical proximity but happen to be in different states from forming a "group." To examine whether this restriction has a material impact on our analysis, we use a density-based spatial clustering algorithm (DBSCAN) to create alternative groups of markets based on markets' geographic proximity to one another (Daszykowski et al. 2002). The clustering algorithm divides markets into density-reachable clusters: market i is directly density reachable from market j if it is within a distance d from market j and market j has at least $k-1$ other markets that are within distance

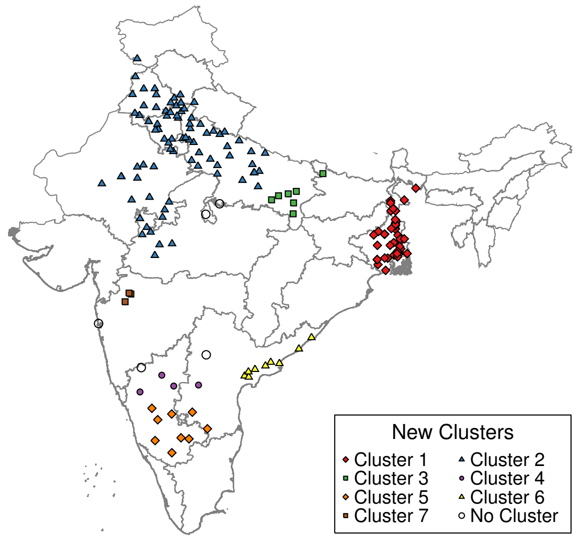


Figure 3 **Comparison of state-based and density-based clusters**

d. The parameters d and k are user-specified inputs. A cluster is a collection of markets that are density reachable either directly or indirectly (i.e., via a series of intermediate directly density-reachable markets). If a market is not directly density reachable from any other market, it is not in a cluster and will not be included in any crop–market-group. This is equivalent to excluding any crop–market-groups that only have one market – a practice followed in the main analysis.

To account for the different trading patterns of each crop, as in the main analysis, we allow the algorithm to potentially form different clusters for each individual crop, i.e., any two markets i and j may be in the same cluster for crop c but in a different cluster for crop c'. For computational tractability we choose $k = 2$ and rather than selecting an arbitrary distance d_c for each of the 159 crops, we set the distance d_c equal to the radius of the circle that would contain $k = 2$ markets if the m markets trading crop c were uniformly distributed in a two-dimensional Euclidean space. Figure 3 shows a map of India with the alternative clusters of markets for a single variety of onion. In this example, $d_c = 89.1$ miles and five markets do not belong to any cluster. A crop–market-group in this specification is a group of markets that are trading a crop within a cluster on a given day. As can be seen in Column 5 of Table 6, the results are qualitatively unchanged when using these new crop–market-groups.

Fourth, we repeat our analysis using a number of different assumptions as outlined below. The results are qualitatively similar to the main results and for brevity are not reported here. (1) We estimate the model using simpler one-way clustered errors (clustered on the crop) and an alternative two-way clustering at the crop and market-group levels. This last specification allows for errors associated with crops that trade within the same market-group to be correlated across time. (2) We confirm that the results are not sensitive to the time window for which we have run the regressions.

We do so by repeating the analysis for the time periods covering August 15, 2010 to November 15, 2010 – an additional week before and after the ban, September 1, 2010 to October 31, 2010 – approximately one less week before and after the ban, and September 11, 2010 to October 16, 2010 – making the before-, during-, and after-ban periods 12 days each. (3) We increase: (a) the number of observations required for a crop–market to be included in the analysis from the 10 used in the main analysis to 15 or 20; (b) the number of markets required in a crop–market-group to be included in the analysis from the two used in the main analysis to three or five; (c) the number of observations per ban stage required for a crop–market-group to be included in the analysis from one, used in the main analysis, to two.

Fifth, the penetration of the RML service is measured with error. In particular, all three penetration measures described in §6.2 are based on the number of subscribers per volume transacted in a crop–market-group and are effectively assuming that this is a good proxy of how influential RML subscribers are in the crop–market-group. This would indeed be the case if the volume transacted by RML subscribers was homogeneous across crop–market-groups for each crop. If not, if for example there were crop–market-groups that had a relatively low (large) number of subscribers who happened to transact a relatively large (small) volume compared to other market participants, this would lead to a classification error – some crop–market-groups that are classified as being in the top quartile of penetration may, in reality, be in a lower quartile and vice versa. Since Model (2) is a multivariate regression, even if the measurement error is classical (i.e., independent and identically distributed), it is not possible to analytically determine the direction of the bias. Nevertheless, the results of a data-driven simulation (available from the authors) suggest that in this case the direction of the bias is toward zero, i.e., measurement error causes attenuation bias. However, if the measurement is not classical, the results of this paper may be affected in a less benign manner.

In the same vein, the measures NSV and SR that were constructed in order to allow for cross-crop comparisons also suffer from different types of measurement error which are worth discussing here. Due to the normalization, the NSV measure will classify all crop–market-groups that belong to crops that have consistently high (or low) penetration across all crop–market-groups as medium-penetration irrespective of their true penetration. The measure SR assumes that there exists a linear relationship between the number of farmers-per-volume-transacted and the price of a crop. Although there is some evidence to suggest that this is indeed the case on average (see footnote 7), there might be crop–market-groups for which this relationship does not hold, resulting in classification errors. In the absence of more comprehensive data at the crop–market-group level, which, as explained in §6.2, is not available, it is difficult to assess how impactful these classification errors are, and in what direction they would influence the results of the paper. We note this is one limitation of this work.

Table 7 Speed of price convergence descriptive statistics

	Overall		**Pre Ban**		**During Ban**		**Post Ban**	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Variability of price differential†								
Overall	0.407	0.680	0.405	0.657	0.423	0.716	0.402	0.703
Quartiles-1–3	0.449	0.738	0.446	0.711	0.364	0.513	0.459	0.773
Quartile-1	0.401	0.325	0.411	0.335	0.409	0.359	0.406	0.362
Quartile-4	0.317	0.525	0.319	0.515	0.517	0.944	0.289	0.524
Mean absolute price differential‡								
Overall	0.143	0.139	0.131	0.138	0.119	0.156	0.120	0.126
Quartiles-1–3	0.155	0.145	0.140	0.139	0.122	0.151	0.136	0.133
Quartile-1	0.197	0.098	0.191	0.103	0.158	0.119	0.177	0.101
Quartile-4	0.117	0.123	0.110	0.134	0.112	0.164	0.088	0.103

†Variability of price differential is the standard deviation of $Q_{cmt} = \ln(P_{cmt}/\bar{P}_{ct})$ over time. ‡Mean absolute price differential is the mean of the absolute deviation $Q_{cmt} = |\ln(P_{cmt}/\bar{P}_{ct})|$ over time. This table summarizes the underlying crop–market–day data that was used to create the crop–market-group-day data of Table 2

7. Speed of Price Convergence

In the previous section we used the ban to show that the information provided by RML is responsible for reducing geographic price dispersion by approximately 12% on average. In this section we complement this analysis by investigating how prices change over time. In general, local shocks in supply and demand may cause the price of a crop in any one market to deviate from the average price. One would expect that such deviations would be short-lived: market participants would adjust their selling and purchasing decisions accordingly (e.g., by shifting supply (demand) away from markets that have a lower (higher) than average price), pushing the price towards the average. One mechanism by which market participants become informed about price deviations in nearby markets is through the information provided by RML. In this section we investigate whether the price adjustment process, and in particular the speed with which prices converge towards the average price, is altered in any way when the RML information is no longer available.

To examine whether prices are indeed converging towards an average price we adopt a set of measures and tests developed in the macroeconomics literature to investigate the law of one price empirically (see Parsley and Wei 1996, Fan and Wei 2006). Note that in order to examine how prices change over time, it is necessary to change the unit of analysis from the crop–market-group–day to the crop–market–day level. However, to keep the analysis of this section comparable with that in §6, and in particular to have a well-defined measure of relative penetration that takes into account which other markets are open concurrently with the focal market (i.e., the crop–market-group), we focus only on those crop–markets that were included in the analysis in §6. This specification leads to a panel containing 2,441 crop–markets over 75 days. As with the crop–market-group–day panel this is an unbalanced panel.

We proceed by following Fan and Wei (2006) to define the price difference measure Q_{cmt} as the percentage difference in price of crop c in market m on date t from the average price of that crop on that day, i.e., $Q_{cmt} = \ln(P_{cmt}/\bar{P}_{ct})$, where the average $\bar{P}_{ct}$ is taken over all markets m trading crop c on day t, which exists for every day t. Naturally, if the law of one price prevails, then the price difference Q_{cmt} should be close to zero (it may not be exactly zero due to transportation costs). A positive (negative) Q_{cmt} signifies that the price of crop c in market m at time t is higher (lower) than the average price. As can be seen from the summary statistics in Table 7, there is substantial price variation over time as the standard deviation of the price difference Q_{cmt}, as well as the mean absolute price difference $|Q_{cmt}|$, is statistically different than zero. Further, the crop–markets that belong to highest top quartile of RML penetration exhibit lower overall price dispersion than the crop–markets in the lower penetration quartiles.

To examine whether prices are converging over time, we first test whether we can reject the hypothesis that the price difference Q_{cmt} follows a random walk (i.e., a non-stationary process). To do so we use the panel version of the Dickey–Fuller framework (Dickey and Fuller 1979, Levin et al. 2002). Namely, we estimate an equation of the form

$$\Delta Q_{cmt} = \delta_t + \alpha_{cm} + \gamma_{ckt} + \gamma_{ckt-1} + \lambda_1 Q_{cmt-1} + \epsilon_{cmt}, \tag{3}$$

where Δ is the first difference operator and δ_t are date fixed effects, α_{cm} are crop–market fixed effects, and γ_{ckt} and γ_{ckt-1} are current- and previous-period crop–market-group fixed effects, respectively. In this specification we include the crop–market fixed effects α_{cm} to control for any systematic differences in prices and speed of convergence across crop–markets. Further, prices and, subsequently, the speed of convergence could be affected by other markets that happen to be trading crop c on a given day t. Since the model compares relative price changes between periods t and $t-1$, we include fixed-effect controls for the crop–market-groups in both periods t and $t-1$.

If the coefficient λ_1 is negative and significantly different from zero, then we can reject the unit root hypothesis and conclude that the prices converge over time. More specifically, a negative λ_1 implies that random shocks at time t, which push the price of crop c in market m away from the average price, are on average followed at time $t+1$ by shocks in the opposite direction. Thus, prices will tend not to wander away from their average price by too much or for too long. However, establishing whether the coefficient λ_1 is significantly different from zero is not straightforward as under the null hypothesis of a unit root, standard errors do not follow the standard distribution, even asymptotically (Dickey and Fuller 1979). To check whether prices are stationary, we perform a Fisher-type test, based on the augmented Dickey–Fuller (ADF) test, as suggested by Maddala and Wu (1999). This test can handle the unbalanced nature of our panel. Besides being unbalanced, a

Table 8 Speed of price convergence

VARIABLES	(1) ΔQ_{cmt}	(2) ΔQ_{cmt}	(3) ΔQ_{cmt}	(4) ΔQ_{cmt}	(5) ΔQ_{cmt}	(6) ΔQ_{cmt}
Q_{cmt-1}	-0.637***	-0.635***	-0.628***	-0.661***	-0.656***	-0.652***
	(0.031)	(0.035)	(0.040)	(0.042)	(0.042)	(0.043)
$DuringBan \times Q_{cmt-1}$		0.003	-0.020		0.010	-0.063*
		(0.018)	(0.013)		(0.034)	(0.038)
$PostBan \times Q_{cmt-1}$		-0.005	-0.006		-0.016	-0.014
		(0.016)	(0.019)		(0.019)	(0.018)
$Quartile\text{-}4 \times Q_{cmt-1}$			-0.030			-0.004
			(0.063)			(0.062)
$DuringBan \times Quartile\text{-}4$			-0.012			-0.031
			(0.009)			(0.019)
$PostBan \times Quartile\text{-}4$			-0.013			-0.025
			(0.011)			(0.027)
$DuringBan \times Quartile\text{-}4 \times Q_{cmt-1}$			0.073**			0.100***
			(0.035)			(0.027)
$PostBan \times Quartile\text{-}4 \times Q_{cmt-1}$			0.005			-0.003
			(0.022)			(0.017)
Observations	77,918	77,918	77,918	42,215	42,215	42,215
Adjusted R-Squared	0.350	0.351	0.352	0.360	0.361	0.363
No. of crop–markets	2441	2441	2441	1861	1861	1861

*** p<0.01, ** p<0.05, * p<0.1, crop–market–day panel regressions. Standard errors, with two-way clustering at the crop level and at the date level, reported in parentheses. Δ denotes the first difference operation and $Q_{cmt} = \ln(P_{cmt}/\bar{P}_{ct})$. All columns include crop–market fixed effects, date fixed effects, and current-period and previous-period crop–market-group fixed effects. Columns 1–3 use all data and lump together the first three quartiles. In columns 4–6 data from the second and third quartile of penetration are excluded.

further complication arising from the fact that not all markets are open every day, is that the time interval between observations may not be constant. We account for this by "closing up" the gaps in the series, which, as shown by Ryan and Giles (1998), preserves the asymptotic distribution of the ADF test statistic. The null hypothesis is that all of the price time series that make up our panel are non-stationary (i.e., contain a unit root), while the alternative hypothesis is that at least one is stationary. We run a number of alternative Fisher-type tests, allowing for crop–market-specific intercepts, crop–market-specific time trends, and autoregressive terms. In all cases the test rejects the non-stationarity hypothesis (p-value<0.1%). Therefore, we proceed under the assumption that the price data is stationary and use standard distributional assumptions for inference.

The results of the estimation of Model (3) are presented in Column 1 of Table 8. The negative coefficient (λ_1=-0.637, p-value$<$ 0.1%) of Q_{cmt-1} demonstrates that the prices of agricultural crops converge, i.e., fluctuations from the national average attenuate over time. Nonetheless, the same result suggests that, to some extent, there is intertemporal predictability in prices, i.e., fluctuations from the national average do not dissipate immediately. In fact, the magnitude of λ_1 implies that random deviations from the mean have a half-life of $\ln(0.5)/\ln(1+\lambda_1) = \ln(0.5)/\ln(1-0.637) = 0.685$ periods (95% C.I. 0.570–0.800 periods).

After demonstrating that prices are converging, we examine the impact of the ban on the speed of convergence. As a first step, we do so by interacting the previous day's price difference Q_{cmt-1} with the ban variables ($DuringBan_t$ and $PostBan_t$):

$$\Delta Q_{cmt} = \delta_t + \alpha_{cm} + \gamma_{ckt} + \gamma_{ckt-1} + \lambda_1 Q_{cmt-1} + \lambda_2 DuringBan_t \times Q_{cmt-1} + \lambda_3 PostBan_t \times Q_{cmt-1} + \epsilon_{cmt}. \tag{4}$$

Column 2 of Table 8 shows the results of Model (4). The speed of convergence is slightly slower during the ban (as indicated by the positive $\lambda_2 = 0.003$ coefficient) but not significantly so (p-value=89.1%), suggesting that the average impact of the absence of RML information on the speed of convergence is not different from zero. Similarly, after the ban is lifted the speed of converge is again very similar to pre-ban levels (as indicated by the negative $\lambda_4 = -0.005$, p-value=77.3%).

We focus next on those markets for which RML has relatively high penetration. To do so, we estimate a model in which the ban is interacted with the quartiles of the relative penetration variable and the price difference of the previous period Q_{cmt-1}. This model specification requires a triple interaction. To avoid having to estimate a model that is too complicated to interpret, we (i) lump together the first three quartiles of the penetration variable and (ii) drop the second and third quartiles from the analysis. (For comparison purposes, the equivalent results for the geographic price dispersion appear in Columns 6 and 7 of Table 6.) More specifically, we estimate

$$\begin{aligned}\Delta Q_{cmt} = {} & \delta_t + \alpha_{cm} + \gamma_{ckt} + \gamma_{ckt-1} + \lambda_1 Q_{cmt-1} + \lambda_2 DuringBan_t \times Q_{cmt-1} + \lambda_3 PostBan_t \times Q_{cmt-1} \\ & + \lambda_4 Quartile\text{-}4_{ck} \times Q_{cmt-1} + \lambda_5 DuringBan_t \times Quartile\text{-}4_{ck} + \lambda_6 PostBan_t \times Quartile\text{-}4_{ck} \\ & + \lambda_7 DuringBan_t \times Quartile\text{-}4_{ck} \times Q_{cmt-1} + \lambda_8 PostBan \times Quartile\text{-}4_{ck} \times Q_{cmt-1} + \epsilon_{cmt},\end{aligned} \tag{5}$$

where all variables are as defined above. An advantage of this formulation is that the penetration variable is defined at the crop–market-group level, i.e., a crop–market becomes a "high-penetration" ("low-penetration") crop–market if it is open with a group of markets for which there collectively exist a large (small) number of RML subscribers per volume transacted. If temporal price fluctuations converge slower during the ban in those areas with relatively high RML penetration, we would expect to see a positive coefficient λ_7 of the triple interaction during the ban. Column 3 of Table 8 shows the results of Model (5), where we use all of the data and lump together the first three quartiles. The coefficient of the triple interaction is positive during the ban ($\lambda_7 = 0.073$, p-value=3.7%) but is not statistically different from zero post-ban (λ_8=0.005, p-value=80.9%). Furthermore, the two coefficients are statistically different from each other (p-value=8.2%). The results of Models (3)–(5), where we drop data from the second and third quartiles, appear in Columns 4–6 of Table 8 and are qualitatively similar. This suggests that prices are indeed converging slower when the information provided by RML is not present, but only in those markets in which RML has a high level of penetration. In particular, the increase in the half-life of random deviations from the mean in the high-penetration markets during vs. before the ban is 0.125 periods larger than the same difference for low-penetration markets. This is equivalent to an 18.3% decrease in the speed of convergence relative to before the ban. When the ban is lifted, the speed of convergence returns to pre-ban levels. These results demonstrates that RML has an impact not only on aggregate geographic price dispersion, but also on the speed at which market prices converge over time.

8. Conclusions

New ICTs such as mobile phone networks are rapidly changing supply chains in developing economies by reducing the cost of acquiring information. Previous research has shown that access to this new technology enables farmers to strategically choose markets in which to sell their produce, correcting demand–supply mismatches and thus reducing geographic price dispersion. In this paper we examine how the provision of regular, timely, and accurate price information delivered via existing ICT impacts geographic price dispersion over and above the impact of having access to the ICT itself. We show that this information has the potential to reduce geographic price dispersion by 12% on average, over and above having access to the ICT itself. Furthermore, we show that it can speed up the attenuation of any random fluctuation from the average price by 18.3%.

The natural experiment identification strategy, combined with a difference-in-differences specification and a detailed set of spatial and temporal fixed effects, greatly reduces concerns regarding endogeneity and/or unobserved heterogeneity. We also explore several alternative explanations, as well as potential issues of misspecification and spurious correlation, and find no evidence that casts doubt on the causal nature of the results.

Our work has managerial and policy implications. Managerially, this study provides support for the business model of third-party information providers, such as RML, by showing that they make a significant difference to the functioning of agricultural produce markets in the developing world. It is therefore understandable that this business model is proliferating.[9] Policy makers and

[9] A number of initiatives that aim to provide rural farmers and other market participants with price and advisory information, such as RML, are being introduced in both India and other developing countries. Examples in India include IKSL (a partnership between the Indian Farmers Fertilisers Cooperatives and Bharti Airtel, an Indian mobile operator), Fisher Friend (a program funded by the M.S. Swaminathan Research Foundation in partnership with Qualcomm, an international technology company, and Tata Teleservices, an Indian mobile operator), and Nokia Life tools (a service offered by Nokia, a European handset maker) (Mittal et al. 2010). Examples in Western Africa include Esoco, a private for-profit company that receives funding from USAID (see USAID 2010), and Manobi, another private for-profit company that works in partnership with Sonatel, a mobile phone operator (see `http://www.manobi.net`, accessed July 29, 2014). Other third-party initiatives that combine information services with a platform that facilitates exchange include Google Trader, a partnership between Google, the Internet search provider, and MTN, a mobile phone operator, which operates in Uganda and Ghana (`http://www.google.com/local/trader/`, accessed July 29, 2014), and Ekhanei in Bangladesh (`http://www.ekhanei.com/`, accessed July 29, 2014).

international aid organizations should take our results into consideration when deciding how to allocate funds aimed at improving welfare in developing countries. As reduction in price dispersion is associated with economically significant improvements in social welfare (e.g., see Jensen (2007)), complementing ICT infrastructure projects with subsidies to farmers for the purchase of price information, such as that provided by RML, or providing support to organizations offering such services, could be an option for further reducing market inefficiencies and improving welfare. Importantly, there are two reasons to believe that for-profit companies may underprovide such services. First, there is evidence to suggest that farmers share the information they receive from services such as RML (e.g., Fafchamps and Minten (2012) find that farmers who receive RML information are likely to share it) and second, our results suggest that there is an informational externality, (i.e., once a critical threshold of penetration has been reached, farmers will receive the benefits of lower price dispersion without having to subscribe to the service). Therefore, for-profit companies may find it difficult to reap the full economic benefit of providing this type of service.

Our rich data set, coupled with the natural experiment specification, has allowed for a robust investigation of the value of timely and accurate information. Nonetheless, there are still a number of limitations. First, while we show that crop–market-groups with a relatively high level of penetration are affected by RML price information over and above the fluctuations in price dispersion observed for crop–market-groups with low penetration, we do not have a control group that is not affected at all by the information provided by RML. Furthermore, the fact that penetration is measured with error makes the results vulnerable to measurement-error bias.

Second, the bulk text message ban was limited in length. This limits the statistical power for examining price dispersion changes; for example, it does not allow us to tease out whether the information is more useful for some types of crops than others (e.g., perishables vs. storable crops, cash vs. staple crops, etc.) and only allows us to assess the impact of RML under a "partial equilibrium" adjustment, i.e., the short duration of the ban may not have given farmers sufficient time to respond by adjusting their "information-gathering" behavior. Nevertheless, there are three reasons to believe that the only substantive information that RML subscribers would have not had access to during the ban is the information provided by RML, suggesting that the long-term impact of RML is likely to be similar to the short-term impact estimated in our paper: (1) The primary channel through which farmers collect information in the absence of RML is their network of friends and family. These networks tend to be fairly cohesive and would have been readily available to RML subscribers during the ban; (2) Most RML subscribers were relatively new to the service at the time of the ban (e.g., 51.5% (95.1%) of subscribers had joined RML less than 6 (12) months before the ban) making it unlikely that they would have completely severed links with their pre-RML information-providing networks; (3) The high incidence of disruptions to mobile phone technology in India makes it likely that RML subscribers would have needed to maintain their informational networks as a back-up for those occasions when RML messages failed to deliver.[10]

Third, the nature of our data does not allow us to pinpoint whether the information provided by third parties such as RML is more useful to the supply side (i.e., the farmers) or the demand side (i.e., the traders/wholesalers/consumers). Nevertheless, the reduction in price dispersion caused by this information should be of benefit to both the demand and the supply side.

Appendix. Estimating Market Penetration

Let the set $\mathcal{S}_{ck}$ denote the set of markets trading crop c that belong to market-group k and let G_c denote the number of market-groups k that trade crop c. Also, let the set $\mathcal{T}_{ck}$ denote the dates that crop–market-group ck is open. The number of markets in crop–market-group ck is then given by $M_{ck} := |\mathcal{S}_{ck}|$ and the number of days market-group ck trades is given by $N_{ck} := |\mathcal{T}_{ck}|$. We also define S_{mt} as the number of RML subscribers that receive information about market m on day t and V_{mt} (P_{mt}) as the volume (price) transacted in market m on day t. The average number of subscribers and volume transacted at crop–market-group ck are given by $n_{ck} = \frac{\sum_{t\in\mathcal{T}_{ck}}\sum_{m\in\mathcal{S}_{ck}} S_{mt}}{N_{ck}}$ and $TV_{ck} = \sum_{t\in\mathcal{T}_{ck}}\sum_{m\in\mathcal{S}_{ck}} V_{mt}$, respectively. The number of subscribers per volume is $SV_{ck} = n_{ck}/TV_{ck}$, the normalized number of subscribers per volume is $NSV_{ck} = SV_{ck}/AV_c$, where $AV_c = \frac{\sum_k (n_{ck}/TV_{ck})}{G_c}$, and the number of subscribers per rupee is $SR_{ck} = \frac{n_{ck}}{TV_{ck}P_c}$, where P_c is the volume-weighted average price of crop c, i.e., $P_c = \frac{\sum_k\sum_{t\in\mathcal{T}_{ck}}\sum_{m\in\mathcal{S}_{ck}} V_{mt}P_{mt}}{\sum_k\sum_{t\in\mathcal{T}_{ck}}\sum_{m\in\mathcal{S}_{ck}} V_{mt}}$.

References

Aker, J. C. 2008. Does digital divide or provide? The impact of cell phones on grain markets in Niger. *Center for Global Development Working Paper Number 154* 1–64.

Aker, J. C. 2010. Information from markets near and far: Mobile phones and agricultural markets in Niger. *American Economic Journal: Applied Economics* **2**(July) 46–59.

Aker, J. C., I. M. Mbiti. 2010. Mobile phones and economic development in Africa. *Journal of Economic Perspectives* **24**(3) 207–232.

Bakos, J. Y. 1991. Strategic analysis of electronic marketplaces. *MIS Quarterly* **15**(3) 295–310.

Bartsch, U., A. S. Gupta, M. Sharma. 2010. India economic update. *World Bank Report* 1–25.

Baye, M. R., J. Morgan. 2001. Information gatekeepers on the Internet and the competitiveness of homogeneous product markets. *The American Economic Review* **91**(3) 454–474.

Baye, M. R., J. Morgan, P. Scholten. 2004. Price dispersion in the small and in the large: Evidence from an Internet price comparison site. *Journal of Industrial Economics* **52**(4) 463–496.

Baye, M. R., J. Morgan, P. Scholten. 2006. Information, search, and price dispersion. T. Hendershott, ed., *Economics and Information Systems*, chap. 6. Elsevier, Oxford, UK, 323–376.

[10] See for example these widely reported disruptions: `http://indianexpress.com/article/india/india-others/telcos-file-complain-against-disruption-of-services/`, accessed July 29, 2014.

Brynjolfsson, E., M. D. Smith. 2000. Frictionless commerce? A comparison of Internet and conventional retailers. *Management Science* **46**(4) 563–585.

Cameron, A. C., J. B. Gelbach, D. L. Miller. 2011. Robust inference with multiway clustering. *Journal of Business & Economic Statistics* **29**(2) 238–249.

Chen, S., M. Ravallion. 2008. The developing world is poorer than we thought, but no less successful in the fight against poverty. *World Bank Report* 1–44.

CIA. 2012. *The CIA world factbook 2012*. Skyhorse Publishing, New York, NY.

Daszykowski, M., B. Walczak, D. L. Massart. 2002. Looking for natural patterns in analytical data. 2. Tracing local density with OPTICS. *Journal of Chem. Infor. and Comp. Sci.* **42**(3) 500–507.

Devalkar, S. K., R. Anupindi, A. Sinha. 2011. Integrated optimization of procurement, processing, and trade of commodities. *Operations Research* **59**(6) 1369–1381.

Dickey, D. A., W. A. Fuller. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**(366) 427–431.

Einav, L., T. Kuchler, J. Levin, N. Sundaresan. 2015. Assessing sale strategies in online markets using matched listings. *American Economic Journal: Microeconomics* **7**(2) 215–247.

Ellison, G., S. F. Ellison. 2009. Search, obfuscation, and price elasticities on the Internet. *Econometrica* **77**(2) 427–452.

Fafchamps, M., B. Minten. 2012. Impact of SMS-based agricultural information on Indian farmers. *The World Bank Economic Review* **26**(3) 383–414.

Fan, C. S., X. Wei. 2006. The law of one price: Evidence from the transitional economy of China. *The Review of Economics and Statistics* **88**(4) 682–697.

Futch, M., C. McIntosh. 2009. Tracking the introduction of the village phone product in Rwanda. *Information Technologies & International Development* **5**(3) 54–81.

Gatti, J. R. J., P. Kattuman. 2003. Online price dispersion within and between seven European countries. M. R. Baye, ed., *Advances in Applied Microeconomics: Vol. 12*, chap. 5. JAI Press, 107–141.

Ghose, A., Y. Yao. 2011. Using transaction prices to re-examine price dispersion in electronic markets. *Information Systems Research* **22**(2) 269–288.

Goyal, A. 2010. Information, direct access to farmers, and rural market performance in central India. *American Economic Journal: Applied Economics* **2**(3) 22–45.

Hsiao, C. 2003. *Analysis of Panel Data*. 2nd ed. Cambridge University Press.

IAMAI. 2012. Internet in Rural India. *Internet and Mobile Associtation of India Report* 1–12.

Isard, P. 1977. How far can we push the "Law of one price"? *The American Economic Review* **67**(5) 942–948.

Jensen, R. 2007. The digital provide: Information (technology), market performance, and welfare in the south Indian fisheries sector. *The Quarterly Journal of Economics* **122**(3) 879–924.

Jensen, R. T. 2010. Information, efficiency, and welfare in agricultural markets. *Agricultural Economics* **41** 203–216.

Levin, A., C.-F. Lin, C.-S. J. Chu. 2002. Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics* **108**(1) 1–24.

Maddala, G. S., S. Wu. 1999. A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics* **61**(S1) 631–652.

Malone, T. W., J. Yates, R. I. Benjamin. 1987. Electronic markets and electronic hierarchies. *Communications of the ACM* **30**(6) 484–497.

Markides, C. 2009. Reuters Market Light: A new business model in action. *London Business School, Case Study* 1–20.

Mittal, S., S. Gandhi, G. Tripathi. 2010. Socio-economic impact of mobile phones on Indian agriculture. *Indian Council for Research on International Economic Relations Working Paper* **246** 1–53.

Overby, E., C. Forman. 2015. The effect of electronic commerce on geographic purchasing patterns and price dispersion. *Management Science* **61**(2) 431–453.

Parsley, D. C., S.-J. Wei. 1996. Convergence to the law of one price without trade barriers or currency fluctuations. *The Quarterly Journal of Economics* **111**(4) 1211–1236.

Ryan, K. F., D. E. Giles. 1998. Testing for unit roots in economic time-series with missing observations. T. Fomby, R. Hill, eds., *Advances in Econometrics*. JAI Press, 203–242.

Sorensen, A. T. 2000. Equilibrium price dispersion in retail markets for prescription drugs. *The Journal of Political Economy* **108**(4) 833–850.

Svensson, J., D. Yanagizawa. 2009. Getting prices right: The impact of the market information service in Uganda. *Journal of the European Economic Association* **7**(2-3) 435–445.

Tambe, P., L. M. Hitt. 2013. Measuring information technology spillovers. *Information Systems Research* **25**(1) 53–71.

Tang, Z., M. D. Smith, A. Montgomery. 2010. The impact of shopbot use on prices and price dispersion: Evidence from online book retailing. *International Journal of Industrial Organization* **28**(6) 579–590.

Thomas, S. 2003. Agricultural commodity markets in India: Policy issues for growth. *Working Paper* 1–25.

USAID. 2010. Using ICT to provide agriculture market price information in Africa. *Briefing Paper* (November) 1–5.

Wooldridge, J. M. 2006. Cluster-sample methods in applied econometrics: An extended analysis. *Working Paper* 1–57.

World Bank. 2011. An evaluation of World Bank Group activities in information and communication technologies: Capturing technology for development. *World Bank Report* 1–158.