VOL. 52 NO. 2 (NOV, 2015), 117–148

LAJE LATIN AMERICAN JOURNAL OF ECONOMICS

WORKING AND CARING: THE SIMULTANEOUS DECISION OF LABOR FORCE PARTICIPATION AND INFORMAL ELDERLY AND CHILD SUPPORT ACTIVITIES IN MEXICO*

Edwin van Gameren**
Durfari Velandia Naranjo***

We analyze factors determining women's decisions to participate in the labor market and provide elderly care and nonfinancial support to their (grand)children. We use data from the Mexican Health and Aging Study, a survey of people aged 50 and over, applying a three-equation, reduced-form SUR model. Results suggest that care needs are the driving force behind caregiving activities. Traditional roles also appear to be relevant in the labor force participation decision: women with a closer labor market connection when they were young are more likely to work. Simulations of demographic changes illustrate potential effects for future caregiving and participation rates.

JEL classification: J13, J14, J22, D13

Keywords: Labor force participation; elderly care; care for (grand)children; multigenerational households; seemingly unrelated regression model

1. Introduction

Labor force participation and caregiving activities are competing for the scarce time of many people. In some age categories, two different kinds of caregiving activities are required simultaneously: care provided to aging parents who begin to suffer from functional limitations and care for the individual's own children (Rubin and White-Means, 2009). Although these are activities of a very different nature, both require time that cannot be spent in other ways, in particular on paid labor. Given that, like many OECD countries, Mexico and other Latin American

* The authors are grateful to the editor and anonymous referees of this journal and to seminar participants at Instituto Nacional de Salud Pública, Fundación Santa Fe de Bogotá, Universidad Militar Nueva Granada, Universidad del Valle, and PROESA/Universidad Icesi for their valuable comments and suggestions. All remaining errors are our own.

** Corresponding author, El Colegio de México, Centro de Estudios Económicos. Mexico City. Address: Camino al Ajusco 20, Pedregal de Santa Teresa, México, D.F., C.P. 10740, Mexico. Tel.: 52-55-54493000 ext. 4087. Email: egameren@colmex.mx.

*** Fundación Santa Fe de Bogotá, Centro de Estudios e Investigación en Salud, Bogotá. Email: durfari@gmail.com.

doi 10.7764/LAJE.52.2.117

countries are experiencing the aging of their population (Burniaux *et al.*, 2004; Zúñiga Herrera, 2004; CISS, 2005), it is important to understand how caregiving activities and labor force participation interrelate and whether expansion of services or support may be needed in the future.

In Mexico, public provision of child care facilities is limited while subsidized long-term or elderly care services are almost non-existent; at the same time, privately paid services are too expensive to be a viable alternative for large sections of the population. There is a private market for home care services, especially for domestic services, but for more specialized (nursing) tasks and residential care the market is small and costs are unaffordable for the large majority of the population, while publicly provided services are virtually absent. In principle, child care services are available through the social security system (for formally employed people) and through the *Programa de Estancias Infantiles* (child care programs for those in informal jobs without access to social security, which is about half of the labor force). However, the access criteria and the number of child care slots imply that actual availability is limited (CONEVAL, 2011).

Care provided informally within the family is therefore an important source of nonfinancial support or help for both the elderly and (grand)children. Mexico is not unique in this sense; a similar situation exists in many other Latin American countries as well as southern Europe (CISS, 2008; Pommer *et al.*, 2007). A tradition of extended families in which several generations live together and share responsibility for household chores further stimulates and facilitates that both care for the elderly and care for children and/or grandchildren are arranged within the household. For the generation in the middle, sometimes called the sandwich generation (Miller, 1981), regardless of whether they reside in the same household as the older and younger generations, strong intra-familial pressure to provide care may affect their opportunities to participate in the labor market and contribute an additional source of income to the household.

In this paper we analyze the factors that determine the decision to participate in the labor market, to provide care for the elderly, and to provide nonfinancial support to (grand)children (that is, spend time helping them with a variety of activities). We use data from the Mexican Health and Aging Study (MHAS), a survey applied in 2001 to people aged 50 and older, including their (younger) partners. This survey contains information on the respondents' living situation, as well as information about their children and their parents. Specifically, the respondents answer questions about the financial and nonfinancial care they provide to their (grand)children and to their parents. We estimate a three-equation, reduced form, seemingly unrelated regression (SUR) model of the three decisions in question.

The results suggest that the need for care is the main force that determines the caregiving activities of women aged 45 to 69, more so than the household's economic situation. Traditional roles also appear to be a relevant issue in the labor force participation decision, as women who had a close connection with the labor market in their younger years are more likely to work, while other income in the household reduces participation. With simulations of conceivable demographic changes in Mexico, such as an aging population due to increased life expectancy and reduced fertility rates, we illustrate potential effects for future caregiving and labor participation rates. A predictable increase in the need for long-term care due to the presence of more elderly parents can be compensated if health improvements are achieved, while an immediate decline in caregiving needs results from a reduction in the number of young children. The simulations suggest that the labor force participation rate is not very sensitive to these demographic changes and does not grow considerably even when the need for caregiving activities decreases.

The next section discusses the literature on informal caregiving activities in relation to labor force participation decisions. Section 3 presents the empirical framework for jointly analyzing the three decisions and introduces the data used. Section 4 presents the estimation results, while Section 5 provides some simulation results that highlight the potential consequences of prospective demographic changes. Section 6 concludes the paper.

2. Literature

In the economics literature, theoretical models of the supply of informal elderly care in combination with the labor force participation decisions of potential caregivers are widely available, as well as models that describe decisions regarding labor force participation and child care usage. However, models that jointly analyze the three decisions are scarce.

Basically, models of supply of informal elderly care describe a trade-off between work, leisure, and informal care, typically from the perspective of the caregiver (Nocera and Zweifel, 1996; Nizalova, 2012; Fevang *et* *al.*, 2008). Taking into account that informal care is usually unpaid, the caregiver must directly derive utility from the activity in order to be willing to provide care to his or her parents. Several motives can be distinguished, including altruism, sense of duty, social norms, reciprocity, bequest, and setting an example for children, but in general the model boils down to some mechanism where the caregiver derives utility from the care given to the care recipient. Often, elderly or long-term care (LTC) bought in the market is considered a component of general consumption; neither privately purchased LTC nor publicly provided or subsidized LTC are modeled explicitly, something that fits well with the Mexican situation.

The empirical LTC literature commonly estimates a reduced-form specification that implies that LTC and labor force participation are explained only by exogenous variables, without a direct interaction between employment and LTC. It avoids the discussion regarding the order of decisions to participate in the labor market and provide care; in general it is not clear a priori which decision comes first or what is the causal relation between the two decisions. Chang and White-Means (1995) estimate a two-decision model where the first one is the decision to work and the second is how many hours to work; however, their sample only contains caregivers and ignores the trade-off between labor and care decisions. Several studies investigate the effect of caregiving on labor force participation: some report negative effects (e.g., Ettner, 1996; Bolin *et al.*, 2008) while others do not find an effect (e.g., Wolf and Soldo, 1994; Meng, 2013). Differences are generated, among other factors, by differences in the amount of care that is provided (Carmichael and Charles, 1998, 2003). Heitmueller (2007) emphasizes that accounting for the endogeneity of LTC in the explanation of labor force participation is essential. Carmichael *et al.* (2010) provide evidence that employment and earnings generate opportunity costs and have a negative effect on the willingness to supply informal elderly care. Using cross-sections, Leigh (2010) finds a large negative effect of LTC on participation, but this effect almost completely disappears with panel data. He explains this difference as caused by incomplete control for individual heterogeneity in the cross-section models: "the kinds of people who provide care tend to have low levels of labor force attachment even before or after they have provided that care."

The general framework for child care demand models is set by the work of Connelly (1992) and Ribar (1995). In these, and in many of the later models and applications (e.g., Michalopoulos and Robins, 2002; Blau and Currie, 2004; Tekin, 2007), the focus is on the labor force participation decision of mothers with young children and their demand for formal (paid or subsidized) child care services. Informal care, for example that provided by grandparents, is sometimes explicitly included as a specific class of care while the mother's own time spent with children is generally considered only implicitly. In contrast with the models for elderly care supply, in the perspective of most child care models it is the optimizing mother who determines her labor supply and demand for grandparental childcare; hence it is not the informal caregiver who takes the decisions, as is the case in elderly care models.

Empirical work on child care decisions often includes both the labor force participation for the complete sample and usage of external child care (by professionals or informal caregivers) conditional upon working. Although the child care decision is observed as being conditional upon being employed, the econometric models used generally take the simultaneity of the decisions into account. This construction is often enforced by the available data, in the sense that in many surveys information about child care decisions is only requested from working mothers. In this paper we model the supply of informal child care provided by women over 45 to their (grand)children, in contrast with the child care literature which generally focuses on demand or use of (formal or informal) child care services by mothers confronted with the decision to work or spend more time with their children.

To bring together the two categories of care and their respective streams in the literature, we need an approach that starts from the same perspective. The natural choice seems to be the perspective of the caregiver, as in the elderly care models, and a focus on the supply of informal care, both LTC provided to the elderly and child care provided to (grand)children. One of the few models that combines child care and elderly care and explains the allocation of parental time between the three activities–labor market, child care, and elderly care–can be found in Giménez-Nadal *et al.* (2007), with empirical applications in Giménez-Nadal *et al.* (2010, 2012). Essentially, they combine the two-generation models demonstrated above for elderly care and child care provision into a three-generation modeling framework where the middle generation maximizes its utility and decides how much care to provide to the older generation (elderly care) and the younger generation (child care). In the theoretical model they emphasize the role of setting an example for children (the "demonstration effect"; see also Cox and Stark (2005) for a similar analysis regarding monetary intergenerational transfers) as a motive for LTC, a factor that is indeed shown in empirical implementations, although a substitution effect is also found: more young children in the household reduces the time spent on elderly care and work in the labor market (Giménez-Nadal *et al.*, 2010, 2012). Other motives such as altruism or prevailing social norms and values may be equally valid or more relevant than setting an example per se, especially in a society like Mexico, where family ties are stronger and extended families are more common than in (Northern) Europe and the United States (OECD, 2012). The basic theoretical framework for providing LTC and child care essentially boils down to the same set of equations as explained before, regardless of the motivation behind it.

3. Empirical strategy

Modeling labor force participation conditional on care decision(s) or modeling care conditional upon work outside the home, which is not uncommon in the literature reviewed in the previous section, hides the inherent simultaneity of the decisions, failing to acknowledge that it is not a priori clear which decision comes first. Given that we observe labor force participation and caregiving activities for all sampled people and therefore are not hindered by non-observability of one or more decisions for a potentially endogenously selected fraction of the sample, in Section 3.1 we describe an empirical framework that simultaneously takes the three decisions into account, followed by a presentation of the data and the relevant variables in Section 3.2.

3.1 Model set-up

We propose a reduced form, seemingly unrelated regression (SUR) model of three equations, describing the propensities of the middle generation (parents) to participate in the labor market, to provide informal care to the older generation (grandparents), and provide care for the younger generation ((grand)children):

$$T_l^* = \gamma_l + X_j\beta_{jl} + \varepsilon_l,$$
$$T_e^* = \gamma_e + X_j\beta_{je} + \varepsilon_e,$$
$$T_h^* = \gamma_h + X_j\beta_{jh} + \varepsilon_h,$$

where the dependent variables of the three equations are latent variables representing the propensities of work (T_l^*), elderly care (T_e^*), and care for (grand)children (T_h^*). What are observed are the binary indicators of whether or not the respondent works (T_l), provides care to her parents (T_e), and provides care to her own children or grandchildren (T_h), where, for $i = l$, e, h, each $T_i = 1$ if $T_i^* > 0$ while $T_i = 0$ if $T_i^* \leq 0$. The three binary decisions yield a trivariate, seemingly unrelated (SUR) probit model. The individual and household characteristics that determine the hours spent on the different activities are encompassed by X_j. We consider a reduced-form model for the three simultaneous decisions, using right-hand side (explanatory) variables that enter into all three equations while the endogenous variables (in particular, the number of hours) do not appear on the right-hand side. By implication, we limit ourselves to the marginal effects of explanatory variables on the outcomes without accounting for the direct effects of labor force participation on caregiving activities, for example, or the effects in the opposite direction; it is not a priori clear which order would be correct and it is also difficult to test for this. In doing so we follow Giménez-Nadal *et al.* (2007, 2010) and jointly estimate the three equations while allowing for correlations between their error terms.

3.2 Data

The data used in this paper are from the Mexican Health and Aging Study, or MHAS (*Estudio Nacional sobre Salud y Envejecimiento en México, ENASEM*)[1]. The MHAS is organized as a panel survey, where the baseline survey (held in 2001) is constructed as a nationally representative sample of about 13 million Mexicans aged 50 and over. The survey contains questions about sociodemographic status (including information on children living outside the household), health status, functional limitations, use of health services and other sources of support, current and previous labor status, sources of income, and assets. Both the heads of the selected households as well as their partners were interviewed, independent of their age, resulting in a total sample size (in 2001) of 15,186 individuals from 9,862 households.

In the analysis we focus on the approximately 6,000 women between 45 and 69 years old in the sample, because this is the age at which

1. The data can be accessed at http://www.mhasweb.org/; see Puig *et al.* (2006) and Wong *et al.* (2007) for more information.

it is more likely that the respondent's parents or in-laws are still alive while the (grand)children may still be young enough to require supervision, and where labor force participation is still an issue.[2] Women are usually the most involved in caregiving activities. Unlike Fevang *et al.* (2012), we include respondents whose parents live in the same household and use dummy variables to capture the possible direct effect of this cohabitation condition on caregiving and labor decisions.

Note that we do not use a sample of only those people who still have at least one parent alive or have children or grandchildren. By presenting unconditional estimations we make it possible to perform simulations that consider the higher survival rates of elderly parents and the lower fertility rates of younger generations. Estimations with samples conditional upon having living parents suggest that our construction adequately captures the relevant effects of other variables; only minor changes are encountered.

Dependent variables

Information on labor status is derived from questions regarding business ownership and salaried jobs. Those who respond positively to the question, "During the last year, did you have a principal [secondary] job?" are considered participants in the labor force, and in addition we consider the respondent a participant if they provide an affirmative answer to the question, "Do you (and/or spouse) own a business or farm?" and if he or she is identified as the (co-)owner through the follow-up question "Who owns this business?" Combining salaried workers and business/farm activities results in a (annualized) labor market participation rate of 45.2% of the women aged 45-69.[3]

Information regarding elderly care activities is derived from the survey question, "In the last 2 years, did you (or your spouse) help your parents with basic personal activities such as dressing, eating or bathing because

2. The survey sample includes people aged 50 and over, but partners of the sampled respondents were also interviewed. Given that many men had younger wives, we have a large number of women aged 45-49 who we decided to include in our analysis, even though they may not be fully representative of the entire population in this age group. Performing the analysis without them results in only minor differences.

3. In an earlier version we used a survey question on labor activities that referred to the week prior to the interview, where "Worked" and "Did not work but had a job" were considered participation. Using that measure, only 27.3% of the women worked; this number compares well to other sources (Juarez, 2010; Murrugarra, 2011). We prefer the definition used in the current version because the time periods for the care indicators and labor are more comparable; the data do not provide information regarding care activities in shorter periods.

of a health problem? Exclude help with household chores, errands, and transportation", and specifically from responses to the subsequent question, "Was this help for at least 1 hour a week, or about 100 hours in the last 2 years?" asked only if the respondent gave a positive reply to the previous question. Both questions are asked only if at least one of the respondent's parents is alive. The formulation of the question does not allow for precise identification of the caregiver. Given that the questions are asked of both the sampled respondent as well as their spouse, we decided to combine the information into a variable that indicates that care is given to either the respondent's parent(s) and/or to the parent(s)-in-law. Furthermore, we assume that if care is given, the female spouse is involved in that care. Research in a variety of countries shows that women carry more of the responsibilities of informal long-term care (Hammer and Neal, 2008; Spillman and Pezzin, 2000; Lilly *et al.*, 2007, Giménez-Nadal *et al.*, 2010), suggesting that our assumption is not very restrictive.

Information on nonfinancial support provided to children and grandchildren is obtained through the survey question "In the last two years, have you (or your spouse) spent at least one hour a week helping your children/their spouses/your grandchildren (or those of your spouse)?" An affirmative response results in a positive value for our indicator for child support activities by the respondent. In contrast with the literature that generally focuses purely on child care provided to young children (aged 0-4 or 0-12), the MHAS survey asks for time spent helping (grand) children without specifying the activities. Hence, it may refer both to general household chores as well as to child care itself.

Table 1 shows the number of observations available for the analysis and the number of working and caregiving women in our sample. Provision of support activities for children or grandchildren is much more common than elderly care–47.9% and 7.8%, respectively–which is not surprising given that the number of children and grandchildren of each respondent is potentially large while there is a natural limit on the number of (living) parents. Among working women caregiving for the elderly is slightly more likely than among nonworking women (8.6 vs. 7.0%) but slightly lower when we look at child support activities (47.2 vs. 48.6%). As a first impression, the respondent's support for children seems to be in conflict with labor force participation, while work and elderly care do not seem to affect each other. Of those who provide support to the elderly, 56.2% also provide support to children. Obviously, the same "care combiners" as a fraction of the child supporters is much lower, 9.1%. Child support activities among those who do not care for the elderly, as well as elderly

Table 1. Dependent variables: employment, LTC, and child care

(Percent)

Care for parents	Working and non-financial assistance for children								
	Not working			Working			Total		
	No	Yes	Total	No	Yes	Total	No	Yes	Total
No			93.0			91.4			92.2
Row-%	52.0	48.0	100	53.8	46.2	100	52.8	47.2	100
Column-%	93.9	92.0		92.9	89.6		93.5	90.9	
Yes			7.0			8.6			7.8
Row-%	44.4	55.6	100	43.2	56.8	100	43.8	56.2	100
Column-%	6.1	8.0		7.1	10.4		6.5	9.1	
Total	51.4	48.6	100	52.8	47.2	100	52.1	47.9	100
Observations			3,308			2,729			6,037
			54.8			45.2			100

Source: Authors' calculations based on the MHAS 2001 survey.

care among those who do not provide nonfinancial support to children is slightly lower–47.2% and 6.5% respectively–indicating that caregiving activities may complement each other.

Explanatory variables

As explanatory variables for the three decisions, we include a set of variables that describe the health and living situation of the respondents' parents, another set that accounts for the presence of children and grandchildren, a block that represents the respondent's labor history, as well as demographic and socioeconomic characteristics. All variables are included in all three equations, as is common in a reduced-form framework, although obviously we should expect that the information regarding parents and (grand)children has the greatest effect on support provided to the elderly and children, respectively; cross-effects are captured through inclusion of the determinants in all equations and through the correlations between them.

With regard to the elder generation, we include information on the parents and in-laws. Both the respondent and the spouse are asked whether the father and the mother are still alive, and if so, whether they are in need of help (using the survey question "Due to a health problem, does your mother [father] need any help with basic personal needs like dressing, eating or bathing?") and whether they can be left alone (based on the survey question "Can your mother [father] be left alone for an hour or more?"; this is recoded such that our variable indicates a more severe problem, i.e., that the parent cannot be left alone). The information is combined into variables that count the number of parents and in-laws that match the respective conditions. The respondents have on average 0.79 living parents or in-laws (Table 2; not shown is that 50% of the respondents have at least one living parent or in-law). The majority of them are in such good health that no help is needed; the average respondent has 0.179 parents/in-laws in need of care, while on average they have 0.099 parents/in-laws who cannot be left alone. For 84.4% of all respondents there are no parents/in-laws who need help, and 91.3% of the respondents have no parents/in-laws who cannot be left alone. Furthermore, we include counts of the number of parents/in-laws living with the respondent (on average, 0.039) and the number of parents/in-laws living alone or with their spouse (on average, 0.294); the reference category is formed by parents living with other children or relatives for at least part of the year.[4]

Table 2. Explanatory variables

	Mean	Std.dev.	Min.	Max.
Parents				
#parents/inlaws alive	0.787	0.963	0	4
#parents/inlaws who need help	0.179	0.445	0	4
#parents/inlaws cannot be alone	0.099	0.335	0	3
#par./inlaws living with respondent	0.039	0.210	0	2
#par./inlaws living alone/spouse	0.294	0.676	0	4
(grand)children				
#nonresident grandchildren	7.899	8.640	0	67
nonresident grandchildren under 18	0.760	0.427	0	1
#(great)grandchild in hh	0.619	1.196	0	13
#hh-members aged 0-4 years	0.225	0.554	0	5
#hh-members aged 5-11 years	0.330	0.719	0	8
#hh-members aged 12-17 years	0.396	0.721	0	5

4. Inclusion of information about the age and education level of elderly parents does not add much explanatory power.

Table 2. (continued)

	Mean	Std.dev.	Min.	Max.
Sociodemographic background				
married/living together	0.690	0.462	0	1
#siblings alive	5.100	2.958	0	21
age	56.521	6.335	45	69
educ.: none (ref.cat.)	0.221	0.415	0	1
educ.: primary	0.544	0.498	0	1
educ.: secondary	0.065	0.247	0	1
educ.: technical/commercial	0.095	0.293	0	1
educ.: preparatory or higher	0.075	0.263	0	1
speaks indigenous language	0.068	0.252	0	1
locality size: over 100000 inhab. (ref.cat.)	0.615	0.487	0	0
locality size: 15000-100000 inhab.	0.149	0.356	0	1
locality size: 2500-15000 inhab.	0.086	0.281	0	1
locality size: less than 2500 inhab.	0.150	0.357	0	1
Socioeconomic background				
non-business assets (*$1mln)	0.330	0.584	-0.595	13.67426
hh nonlabor income (*$1000)	3.775	97.660	-500	7500
spousal labor income (*$1000)	2.421	34.065	-500	2292.222
access to medical services	0.641	0.480	0	1
made pension deposits, 1-10 years	0.045	0.208	0	1
made pension deposits, 10-25 years	0.056	0.231	0	1
made pension deposits, >25 years	0.049	0.216	0	1
Health status				
self-assessed health (0-4)	1.260	0.808	0	4
#problems with ADL (0-6)	0.177	0.700	0	6
#problems with IADL (0-4)	0.089	0.426	0	4
suffers a chronic disease	0.616	0.487	0	1
poor mental health status	0.127	0.333	0	1

Source: Authors' calculations based on the MHAS 2001 survey.

Regarding the younger generations, we include the number of (great-) grandchildren who are living in the same household as the respondent, as well as specific indicators for the number of household members under age 5, between 5 and 11, and between 12 and 17. Because children or grandchildren who do not live in the same household as the respondent may live near enough to receive support and attention, we also include indicators of the number of grandchildren of the respondent's nonresident children, as well as an indicator of whether some of them are less than 18 years old.[5] The respondents have, on average, 7.9 nonresident grandchildren, and 76% of the respondents have nonresident grandchildren aged under 18 (Table 2). The total number of household members aged 0 to 17 is much lower–less than 1–but given the proximity they could influence the decision-making process.

There may be some concern about the exogeneity of the explanatory variables, in particular with regard to the fertility decision reflected in the number of grandchildren–potential parents who know that grandparents will help may be more likely to become parents–and the parental living situation, as the decision to live in the same household may be driven by the elderly parent's or the household's care needs. We argue that fertility decisions are not strongly affected by the availability of grandparents; contraceptives and other birth control measures are less widespread than in many other Western countries, the average age at which mothers have their first child is lower, and the number of children per mother is higher (OECD, 2012). Regarding the living arrangements of the elderly, the proportion of elderly living in multigenerational households is much higher than in the U.S. and Europe (OECD, 2012); Aranibar (2001) reports that around 80% of the elderly live in multigenerational households. Traditions, preferences, and considerations of space and available resources of both the elderly as well as the respondent and her siblings are expected to be much more important than care needs alone; furthermore, there is a tendency not only for the elderly to move in with their adult children but also for adult children to continue living with their parents. Hence, although we acknowledge that endogeneity cannot be ruled out completely, we consider that care opportunities play a minor role in those decisions.

Information about pension fund contributions accounts for the connection with the labor market. On the one hand, past contributions imply a history of formal employment (since most pension contributions are provided through the social security system, which requires contributions based on formal employment), while on the other hand prolonged periods of contributions create the opportunity to retire from the labor market with a pension and spend more time on care activities. Table 2 shows that 4.5% of the respondents have fewer than 10 years of contributions; this is generally an insufficient amount of time to claim a pension. About 5.6%

5. The survey provides information on the number of grandchildren but not how many of them are younger than 18.

report between 10 and 25 years of contributions, which qualifies them for a partial pension. With more than 25 years of contributions (4.9% of the respondents), retirement with the maximum pension is possible. Thus, the majority of respondents do not report any pension contributions, indicating the absence of a labor history in the formal sector. Access to health care services provided by the social security system can also be obtained through the formal employment of the spouse or other family members. Table 2 shows that 64.1% report access to these health care services, which implies, given the much lower (formal) employment rates in our sample, that the majority obtains access through derived rights. Social security institutes also provide child care services; hence, although access rights and actual opportunities differ, the same variable tells us something about the availability of formal child care.

We include sociodemographic variables such as age (56.5 years on average), and whether the respondent is married or living in a consensual union (69%). The respondents have on average 5.1 living siblings. About 22.1% of the respondents have no or incomplete primary education, while 54.4% report (completed) primary education as their highest level. Secondary education (generally attended from ages 12 to 15) is the maximum for 6.5% of the respondents, while 9.5% report having finished professional or technical schooling and 7.5% finished high school (*preparatoria*) or higher education (university).[6] Moreover, 6.8% say that they are able to speak an indigenous language. The majority lives in cities with 100,000 or more inhabitants (61.5%, reference category), and only 15% live in towns with fewer than 2,500 inhabitants.

We include five indicators of respondents' health status.[7] A general, subjective indicator is self-assessed health, measured on a five-point scale (poor, fair, good, very good, or excellent health). Another variable indicates whether the respondent suffers from a chronic disease (hypertension, diabetes, cancer, respiratory illness, heart disease, stroke, or arthritis), while we also include a dummy variable that indicates a serious mental health problem. Furthermore, we include two indicators

6. Construction of potential wages would have allowed the estimation of wage elasticities, as is common in the empirical child care literature (see, for example, Connelly and Kimmel, 2003; Michalopoulos and Robins, 2002; Connelly, 1992; Borra, 2010). We prefer to include age, education, and labor history in the main model, allowing them to have direct effects on the decisions, instead of including them only in the wage equation and assuming that their effects run exclusively through potential wages.

7. van Gameren (2008, 2010) presents evidence for Mexico that health is not endogenous to the explanation of labor force participation. There is little evidence in the literature that health is affected by elderly care activities, the strongest effect being a (negative) one on mental health (Coe and Van Houtven, 2009; Schmitz and Stroka 2013).

that report the respondent's limitations in regard to activities of daily living (ADL, applying a measure based upon the Katz Index (Katz *et al.*, 1970; Shelkey and Wallace, 2012) that counts the number of activities a respondent reports problems with: dressing, walking across a room, bathing/showering, eating, getting into or out of bed, and using the toilet) and instrumental activities of daily living (IADL, which reflects the number of problems with preparing hot meals, shopping, taking medication, and managing money). Table 2 shows the average number of ADL and IADL problems; what is not shown is that 91.0% of the respondents do not report any ADL problems and 94.5% do not have IADL limitations.

Possession of non-business assets (total net value of real estate, investments, savings, stocks, shares and bonds, and private means of transport) says something about the resources available for obtaining private care services and the respondent's need to participate in the labor force. Spousal labor income does the same, and here perhaps a larger effect on participation could be expected because if a spouse is working it implies that he is not available for household chores.[8] The respondents report average assets equal to 329,632 Mexican pesos, but the spread is huge, ranging from debts of 600,000 pesos to properties with a total value of 13.7 million pesos. The average spousal monthly labor income (from salaries, bonuses, and business and farm activities including self-employment) is about 2,421 Mexican pesos (about two times the formal minimum income), but also here the range is wide, from large negative incomes up to huge positive incomes. The latter also holds for the average non-labor income of 3,775 pesos, income that includes retirement and other pensions, transfers from government programs as well as children (remittances), and income from property or assets.

4. Results

The first column of Table 3 shows that the driving force behind the decision to provide care to parents is the situation in which the parents encounter themselves: having one or more parents or in-laws alive is (obviously) a relevant factor (first line), and the subsequent lines

8. Note that we ignore the spouse's labor force participation and caregiving decision. We consider this a minor issue, given that the majority of men aged 50-65 work (van Gameren, 2008). In this we follow the (child) care literature, which generally considers the husband's labor decision exogenous. The earnings capacity of other household members is not included in the analysis; in a preliminary version we included the (log of the) total household consumption, but this did not add explanatory power.

indicate that the health status of the parents is of utmost importance. Parents who need help or who cannot be left alone receive support much more often than parents for whom no health problems are reported. Moreover, if the parents live in the same household as the respondent, it is much more likely that the women take care of them than when they live with the respondent's siblings (the reference category), while parents who live alone or together with their spouse receive even less support. The international evidence also finds that living arrangements are important; when the caregiver and elderly relatives live together it is more probable that the caregiver will quit her job in order to look after those relatives.

The information about the respondent's children and grandchildren does not tell us anything about the care given to parents, but the respondent's socioeconomic situation has some relevance. The household's non-labor income has a positive (though decreasing) effect on women's elderly care activities. Only those who have made contributions to a pension fund for less than 10 years are significantly less likely to provide care, in comparison with those who never contributed (the reference category) and with those who contributed longer. A longer period of contributions implies that some pension can already be claimed, while those with fewer than 10 years of contributions still have a chance to qualify for a pension if they contribute for additional years.

The second column of Table 3 shows the factors that explain the respondent's support activities for her children or grandchildren. As expected, a higher number of young household members increases the probability that children are supported. Children younger than 5 years have the strongest positive effect. For children of primary school age (5 to 12), the positive effect is half the size of the effect for pre-school aged children, while no significant effect is found for children over 12, given the positive effect of the number of (great) grandchildren in the household. It is interesting that the number of nonresident grandchildren has no direct effect on child support activities, in contrast with the number of (great)grandchildren in the household. However, if the respondent has nonresident grandchildren aged under 18, her child support activities are significantly greater.

Also interesting is that, apart from the relevance of young (grand) children, the presence of elderly relatives in need of care has a positive effect on child support activities. Something similar was found by Giménez-Nadal *et al.* (2010), who conclude that the presence of

Table 3. Joint LTC, child care, and labor force participation decision

	[1] Elderly care		[2] Child support		[3] Employment	
Parents						
#parents/inlaws alive	0.259 ***	(0.040)	0.017	(0.028)	0.031	(0.028)
#parents/inlaws who need help	0.901 ***	(0.061)	0.108 **	(0.046)	0.092 **	(0.044)
#parents/inlaws cannot be alone	0.254 ***	(0.072)	-0.135 **	(0.057)	-0.087	(0.055)
#par./inlaws living with respondent	0.817 ***	(0.093)	-0.148 *	(0.081)	0.001	(0.081)
#par./inlaws living alone/spouse	-0.138 ***	(0.050)	-0.043	(0.033)	-0.044	(0.033)
(grand)children						
#nonresident. grandchildren	-0.006	(0.005)	-0.001	(0.002)	-0.001	(0.003)
nonresident grandchildren under 18	0.024	(0.074)	0.515 ***	(0.046)	0.032	(0.045)
#(great)grandchild in hh	-0.050	(0.050)	0.126 ***	(0.030)	0.063 **	(0.028)
#hh-members aged 0-4 years	0.004	(0.079)	0.201 ***	(0.047)	-0.117 ***	(0.043)
#hh-members aged 5-11 years	0.018	(0.060)	0.103 ***	(0.037)	-0.059 *	(0.035)
#hh-members aged 12-17 years	0.044	(0.044)	-0.025	(0.027)	-0.038	(0.026)
Sociodemographic background						
married/living together	-0.156 **	(0.073)	-0.004	(0.040)	0.002	(0.040)
#siblings alive	0.022 **	(0.010)	0.010 *	(0.006)	0.004	(0.006)
age	0.222 **	(0.090)	0.094 *	(0.049)	-0.062	(0.049)
age squared (*100)	-0.208 ***	(0.079)	-0.094 **	(0.042)	0.019	(0.043)
educ.: primary	0.083	(0.082)	0.093 **	(0.045)	0.088 **	(0.044)
educ.: secondary	0.155	(0.134)	0.108	(0.079)	0.284 ***	(0.079)

Table 3. (continued)

	[1]		[2]		[3]	
	Elderly care		Child support		Employment	
educ.: technical/commercial	0.320 ***	(0.121)	0.169 **	(0.074)	0.190 **	(0.074)
educ.: preparatory or higher	0.039	(0.157)	0.321 ***	(0.087)	0.331 ***	(0.088)
speaks indigenous language	-0.165	(0.127)	-0.011	(0.069)	0.008	(0.068)
locality size: 15000-100000 inhab.	0.185 **	(0.080)	-0.076	(0.048)	0.109 **	(0.049)
locality size: 2500-15000 inhab.	0.236 **	(0.102)	-0.261 ***	(0.064)	0.158 **	(0.063)
locality size:less than < 2500 inhab.	0.187 **	(0.095)	-0.260 ***	(0.054)	0.185 ***	(0.054)
Socioeconomic background						
non-business assets (*$1mln)	0.050	(0.082)	-0.036	(0.050)	-0.020	(0.053)
squared non-business assets	-0.007	(0.010)	0.000	(0.006)	0.003	(0.007)
hh nonlabor income (*$10,000)	0.139 *	(0.074)	0.020	(0.012)	-0.074 **	(0.030)
squared hh nonlabor income	-0.021 **	(0.010)	-0.000	(0.000)	0.002 *	(0.001)
spousal labor income (*$10,000)	0.013	(0.020)	-0.005	(0.011)	-0.027 *	(0.016)
squared spousal labor income	-0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
access to medical services	0.090	(0.066)	0.192 ***	(0.039)	-0.150 ***	(0.039)
made pension deposits, 1-10 years	-0.337 **	(0.150)	-0.104	(0.083)	0.338 ***	(0.083)
made pension deposits, 10-25 years	-0.014	(0.120)	-0.093	(0.077)	0.587 ***	(0.078)
made pension deposits, >25 years	0.027	(0.142)	-0.038	(0.086)	0.304 ***	(0.087)

Table 3. (continued)

	[1] Elderly care		[2] Child support		[3] Employment	
Health status						
self-assessed health (0-4)	-0.042	(0.040)	0.008	(0.024)	0.043 *	(0.024)
problem with ADL	0.036	(0.053)	0.068 **	(0.032)	0.008	(0.030)
problem with IADL	-0.146	(0.096)	-0.156 ***	(0.051)	-0.149 ***	(0.051)
suffers a chronic disease	0.124 *	(0.064)	0.078 **	(0.037)	0.027	(0.037)
poor mental health status	0.042	(0.089)	0.139 ***	(0.052)	0.061	(0.052)
Constant	-8.226 ***	(2.523)	-3.136 **	(1.402)	2.585 *	(1.407)
ρ_{12}	0.145 ***	(0.036)				
ρ_{13}	-0.021	(0.036)				
ρ_{23}	-0.019	(0.021)				
Wald χ^2 test [a]	794.0 ***	$p = 0.000$	536.2 ***	$p = 0.000$	482.2 ***	$p = 0.000$
Number of observations	6037					
Log likelihood	-8925.8					

Source: Authors' calculations based on the MHAS 2001 survey.
[a] Wald-tests that all parameters equal zero (except the constant); in each equation a $\chi^2(35)$ distribution applies. Robust standard errors; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

children during elderly care activities increases the time devoted to both elderly and child care as primary activities. It is possible that "economies of scale" may be at work in Mexico, in the sense that the grandchildren are around and cared for while care is also provided to the elderly, but a demonstration effect as suggested by Giménez-Nadal *et al.* (2010, 2012) cannot be discarded. If the elderly cannot be left alone, combining care activities appears to be less feasible, as reflected by the significantly negative parameter estimate. The negative effect of the number of parents/in-laws in the same household on the respondent's child support activities reflects another kind of scale benefit: the co-resident elderly can take care of the younger household member, enabling the respondent to spend time on other activities. We find that spousal and non-labor income, wealth, and pension rights do not explain child support activities. In the child care literature (Blau and Currie, 2004) it is common to find that mothers with higher potential wages work more and spend less time with their children. Under the presumption that higher education comes with higher potential wages, our results appear to contradict the general finding regarding child care, but it is important to consider that we are not (only) looking at mothers' time with their own children but (mainly) at the time grandmothers spend with grandchildren. Moreover, education is not the only indicator of earnings capacity. Access to health care services provided by the social security system has a positive effect on within-family child support activities. A negative effect was expected, because child care services are provided through the same social security institutes; however, access to social security is also an indicator of a better socioeconomic status that may make it feasible to stay at home and provide support activities for children.

The third column of Table 3 presents the respondent's labor force participation decision. Here, the economic situation has a stronger effect. In particular, the existence of other incomes in the household, access to health care services, and previous pension contributions are relevant. Access to the health care services offered by the social security system, possibly acquired through working family members, reduces the likelihood that the respondent is active in the labor market. If, however, the respondent has made contributions to a retirement pension in the past, it is more likely that she is still working. Past contributions tell us something about the labor experience of the respondent, as contributions not only indicate a history of formal employment but also more generally that the respondent has not been exclusively dedicated to household chores all her life. In general, it suggests a strong connection with the labor market and therefore the capacity to earn income. The latter is also determined by the level of education achieved, a variable that is also found to be significant in our analysis. Education in general, and secondary or higher levels of education in particular, raise the probability of working. Furthermore, we find that the household's non-labor income and spousal labor income lead to a significant reduction in the probability that the respondent has a job; this is in line, for instance, with Chang and White-Means (1995), who found non-wage income to be a significant variable affecting negatively caregivers' decision to work outside the home. We find no evidence that household wealth has an effect on the respondent's labor force participation decision.

The presence of needy elderly relatives and grandchildren has a small positive effect on the respondent's labor force participation decision. When parents are in need but can still be left alone, and when there are more grandchildren but not any younger than 12 years, additional care needs are probably relatively minor; perhaps another household member takes care of both the elderly and (older) children, permitting the respondent to spend more time on employment in order to acquire additional resources that can be used for care activities. With relatively limited care needs, scale benefits are easier to achieve and leave time for other activities.

Hence, decisions to support the elder and younger generations are primarily governed by the need factors that we would have expected (and some cross-effects), where traditional roles and expectations probably constitute an important underlying reason, but due to inherent measurement problems they are not fully observable. Economic factors are important for the labor force participation decision, but cross-effects on care decisions cannot be ignored. In addition, we find a variety of effects of other background characteristics that suggest interrelationships between the three decisions. For example, the estimates suggest that the maximum burden of elderly care activities is found among women aged 53.3, while for child support activities the maximum occurs at the age of 50.4, and no pattern is found for paid labor. At higher ages it becomes unlikely that there are still living parents around and there will be fewer young children around. While no health effects on elderly care are found, the respondent's own health is relevant in child support activities and labor decisions. The presence of ADL-problems, which do not impede independence in regard to skills such as food preparation and hence do not hinder "light" child support (including companionship), increases the probability that the respondent provides child support activities. But when the respondent suffers IADL problems–when serious aging-related dysfunction and reduction of independence starts–a reduction in child support activities and labor force participation is found. Those who perceive their health as good are more likely to be active in the labor market.[9] If the respondent has a spouse, less care is provided to the parents, but there are no effects on the other decisions. It seems that for unmarried women it is easier to take up care activities for her parents, perhaps because her siblings consider it the responsibility of their unmarried sisters. A small positive effect on caregiving is found if the respondent has more siblings. In the literature it is more common to find that more siblings imply a reduction of care activities for each of them (Nizalova, 2012) and care for the elderly is replaced by work activities.[10] Perhaps the interaction of elderly care duties with migration decisions, as indicated by Antman (2012), may lead to the unavailability of siblings for caregiving and reduce substitution among the various siblings.

Joint decision-making is also implied by the estimates of the correlations between the unexplained parts of the decisions. In particular, ρ_{12} is significantly positive, suggesting that there are unobserved factors that increase both types of caregiving. A preference for caregiving by the women in our sample, but also traditional role models or habits may be at work. Giménez-Nadal *et al.* (2010, 2012) also found a positive relation between child care and elderly care: When children are present while care is being provided to the elderly, there is a booster effect on the total time devoted to caregiving. Unmeasured factors appear less important to the relationship between elderly care and employment (ρ_{13}) as well as between child support activities and labor force participation (ρ_{23}). Caregiving activities and labor force participation do not seem to affect each other in ways that are not captured by the observed variables.

9. In a model without (subjective) self-assessed health, the role of other health indicators hardly changes. When mental health–perhaps the health indicator that is most susceptible to endogeneity (Schmitz and Stroka, 2013)–is left out of the analysis, the results reported in Table 3 are maintained. Completely eliminating the health information from the model results in only minor changes in the other parameters.

10. This somewhat surprising siblings effect disappears if the parental living situation, with 'living with siblings' as the reference category, is excluded from the analysis. An explanation for the siblings effect we found could be that the other siblings want to share the burden of care with the sibling who houses the parent(s).

4.1 Robustness checks

We performed a variety of robustness checks in subsamples of the data set. Specifically, we ran the analysis separately for respondents living with and without a partner, respectively, because having a partner implies differences in potential care needs as well as different opportunities to share the burden. There are some minor differences in the effects found in the two subsamples, but overall the effects are maintained. The needs effect (having parents who are alive, parents in need of care) on elderly care is much larger among women living alone than those living with a partner. For the latter, on the other hand, the age pattern becomes more pronounced for elderly care and for labor force participation, and the relevance of education increases for all three decisions, while age and education effects almost disappear among single women. Among single women we see a stronger effect on participation of earlier contributions to pension funds. The differences may suggest that single women provide care whenever necessary but are also more inclined to work when possible.

Because caregiving responsibilities may differ among age groups, as demonstrated by the age patterns implied by the estimates in Table 3, for example because older women are traditionally less inclined to work outside the home and are less likely to have living parents, we also ran the analysis separately for the younger women in our sample (aged 45-54) and for the older ones (aged 55-69). Indeed, in the older age group we find hardly any effects due to socioeconomic conditions except the pension contribution history, which may directly translate into a future pension; meanwhile, in the younger group socioeconomic conditions are important and other incomes in the household reduce the participation probability. Also, in the older group care needs are more important for explaining specific care activity than in the younger group; in the latter the cross-care effects disappear, while in both groups care needs have hardly any explanatory power for the labor decision. What is particulary interesting is that education level is particularly relevant in the labor decision of the younger age group.

In addition, we ran separate analyses for urban areas with more than 100,000 inhabitants and semi-urban and rural areas because the availability of external services may be more limited in semi-urban/ rural areas than in large cities; also, family ties may be stronger and attitudes towards non-family care more negative. Nevertheless, the results regarding care needs are fairly similar; an exception is the care provided when parents/in-laws live alone or with a spouse, for which the negative relation reported in Table 3 was found only in urban areas. Apparently, respondents in large cities provide less care to their parents, perhaps because the latter do not live in the same city. The socioeconomic situation of the respondent turns out to be much more relevant for participation decisions in large cities, while it is almost irrelevant in more rural areas. Differences in the labor market structure and the role of migration from peripheral areas to large cities may be behind the different relations.

5. Simulations

In order to give a clearer idea of the implications of the estimation results, we present simulations in which we compare the caregiving and labor decisions of typical persons and household situations. The selected simulations reflect demographic trends that are observed in Mexico and indicate how caregiving and participation decisions at the individual level could change due to these trends. Table 4 shows the results of the simulations, in which we fixed a combination of individual characteristics at specific values and used the observed values for all other characteristics to predict the three outcomes (using the estimates in Table 3) for each respondent and calculate the average probabilities.

The global population aging trend, reflected by a larger share of older people in the population and therefore a higher average age, is also occurring in Mexico (Zúñiga Herrera, 2004). In an aging population, it becomes more likely that people aged between 45 and 69 years old have living parents who either need help and cannot be left alone or who are in good health and do not need care. The larger share of elderly people in the population is accompanied by a prospective reduction in the number of young people. In Table 4 (Panel I) we combine these trends in potential scenarios and present the simulated participation in the three modeled activities.

The simulations in panel I of Table 4 suggest that increased elderly care needs due to more living parents are not necessarily made up for by the interaction of expected health improvements of the parents and smaller numbers of young grandchildren; full compensation is found only if none of the parents need care (scenario C). The smaller number of grandchildren and their older ages drastically reduces the probability that women aged 45-69 are expected to provide child support activities. However, the increase in the labor force participation rate does not

Table 4. Probabilites of LTC, child care, and employment

	Pr[elderly care]		Pr[child support]		Pr[participation]	
Panel I: More and healthier parents, fewer and older grandchildren						
A[a]	0.195	(0.046)	0.764	(0.033)	0.402	(0.037)
B[a]	0.315	(0.036)	0.516	(0.028)	0.527	(0.027)
C[a]	0.091	(0.018)	0.475	(0.028)	0.492	(0.027)
Panel II: As in panel I but for different formal labor histories						
A[a], never contrib. (informal)	0.199	(0.047)	0.768	(0.033)	0.378	(0.037)
B[a], never contrib. (informal)	0.320	(0.036)	0.521	(0.028)	0.504	(0.027)
A[a], 10-25yr with contr. (formal)	0.195	(0.054)	0.740	(0.041)	0.598	(0.047)
B[a], 10-25yr with contr. (formal)	0.315	(0.053)	0.485	(0.039)	0.715	(0.034)

Source: Authors' calculations based on the MHAS 2001 survey.
a. Scenario A: One parent alive, in need of care and cannot be left alone, three grandchildren aged under 5 in the household. Scenario B: Three parents alive, of which only one has care needs while all can be left alone, one grandchild aged 12-17 in the household. Scenario C: Same as scenario B, except that none of the parents have care needs. Standard errors (Delta method) in parenthesis.

match the reduction of caregiving activities, especially in the scenario with more living parents when none of them need care, suggesting that care and work activities do not unambiguously compete for the time of middle-aged Mexican women. Other reasons such as traditions and attitudes regarding work may be expected to be important. It is important to remember that female labor force participation in Mexico is already rather low compared with other OECD and Latin American countries (Arceo and Campos, 2010; van Gameren, 2010), and that with an aging population, the share of working-age people will decline. In such a situation the participation rate should increase substantially to maintain the same production levels, where only a minor increase is suggested by our simulations.

Panel II presents scenarios A and B separately for women who have been more attached to the labor market in earlier years (last two lines, between 10 and 25 years of contributions) and for those who have never been attached to the formal labor market (first two lines, never made pension deposits). Women with a strong connection to the formal labor market are much more likely to continue their participation in the labor market than women who have not worked formally before. The labor market history in itself does not affect incentives for caregiving; the probabilities of providing child support activities and elderly care reported in panel I remain rather similar when calculated separately for the labor market connection, but in combination with the other demographic trends, the observed changes in the labor force participation of Mexican women–nowadays participation among young women is higher than for the generation represented in our data–will have important consequences for the availability of caregivers that may go beyond what we can highlight with our analysis.

Note that we have not imposed restrictions on the feasibility of the simulated outcomes; these are partial results assuming that there are no changes in behavior, the household environment, and other circumstances (a structural model would only partially overcome these drawbacks). The results are indicative of what could happen if there are no other changes. In particular, feasibility will depend on the drastically changing shares of the various age categories in the population; a lower individual propensity of women aged 45-69 to provide elderly care might not be tenable if the number of elderly with care needs increases faster than the number of women aged 45-69. They are more likely to be feasible if the population aging process is accompanied by health improvements, that is, by a growth in the number of years of good health more than by longer life expectancy in itself, because it is less likely that a conflict between elderly care and labor will arise.

6. Conclusions

We have analyzed simultaneous decisions made by Mexican women aged 45 to 69 years regarding labor force participation, caring for parents (known as elderly or long-term care), and providing support activities to (grand)children. A tradition of extended families in which multiple generations live together in combination with limited availability of affordable (public or private) elderly and child care facilities implies a large dependence on informal support activities for the elderly and (grand)children. The estimation results of a reduced form, seemingly unrelated regression (SUR) model confirm that care needs, rather than the economic situation, are the driving force behind caregiving activities. Having parents or in-laws with health problems strongly increases the probability that elderly care is provided, but also has (small) effects on provision of child support activities and labor force participation. Young grandchildren raise the probability that the respondent provides support activities to (grand)children, while it slightly reduces the probability of labor market activities. Hence, care needs not only increase the care activity required, but also have some relevance for other activities.

Economic considerations are the greatest component of the participation decision. Income from other sources reduces labor force participation and has a (smaller) positive effect on the probability that elderly care is provided. Women with a close connection to the formal labor market earlier in their lives are more likely to work, a connection that nonetheless has no effect on caregiving activities, signalling a potential double burden of care and work for those women. The important effect of own earnings capacity on labor force participation is evidenced by the role of educational levels–a higher level of education predicts participation–and women who report better health (fewer health problems) are more likely to be active in the labor market and less likely to perform caregiving activities. In general, traditional roles that prescribe that women provide care when necessary appear to be relevant, but in addition we find several indications of the interdependence of the three decisions analyzed. Furthermore, we find evidence that there are other (unobserved) characteristics that increase both types of caregiving activities: preferences as well as scale effects may be behind the interdependencies.

We have illustrated potential future caregiving and participation rates with simulations of demographic changes in Mexico reflecting the aging population. Although the Mexican population is still rather young compared to many (Southern) European countries, the onset of an aging population is observable, with reductions in both fertility and mortality rates. A scenario with more parents living but in better health than the elderly in our sample, in combination with fewer and older grandchildren, is likely to reduce individual care needs but lead to only a small increase in labor force participation rates. Increased participation rates can be expected if future generations of women have a stronger connection to the (formal) labor market.

Taken together, our results highlight that an expansion of the affordable (public) provision of elderly care and/or child care will not replace informal caregiving on a wide scale, since informal care provision is an embedded habit for many; expansion of affordable services may have relatively small consequences on caregiving and work decisions and therefore on the burden involved. Nevertheless, it is important to be aware that for the scenarios, we assume the absence of behavioral changes among the generation in the middle, something that is unlikely in the long run. Drastically changing shares of age groups in the population, in particular a faster increase in the number of older elderly persons in need of care than in the number of women available for caregiving activities, combined with increases in the labor force participation rates of younger women, may imply that current informal care activities are not tenable and behavioral changes are unavoidable.

That the expansion of public subsidies for or provision of caregiving may lead to the reduction of private or informal care has been shown in various countries. Mexico is not likely to be different; results by Juarez (2009) indicate that the expansion of public pension transfers almost completely crowded out private financial support given to the elderly. Additionally, the availability and accessibility of more services for the elderly and children may have an effect on quality of life and the well-being of those caregivers who feel burdened by their care tasks, especially for those who combine care with paid work or feel forced to consider providing care and working simultaneously. Traditions, attitudes, and opinions about caregiving may change and the consequence may be greater willingness to consider and use external services (public or private). For that to be successful, not only affordability but also quality and safety of the services must be guaranteed.

REFERENCES

Antman, F.M. (2012), "Elderly care and intrafamily resource allocation when children migrate," *Journal of Human Resources* **47**(2): 331–63.

Aranibar, P. (2001), "Acercamiento conceptual a la situación del adulto mayor en América Latina," CEPAL Serie Población y Desarrollo No. 21, Santiago de Chile.

Arceo Gómez, E.O. and R.M. Campos-Vázquez (2010), "Labor supply of married women in Mexico: 1990-2000," El Colegio de México, Centro de Estudios Económicos, Documento de Trabajo No. 2010-16.

Blau, D.M. and J. Currie (2004), "Preschool, daycare, and afterschool care: Who's minding the kids?" National Bureau of Economic Research Working Paper 10670, Cambridge, MA.

Bolin, K., B. Lindgren, and P. Lundborg (2008), "Your next of kin or your own career? Caring and working among the 50+ of Europe," *Journal of Health Economics* **27**: 718–38.

Borra, C. (2010), "Childcare costs and Spanish mothers' labour force participation," Hacienda Pública Española / *Revista de Economía Pública* 3/2010 (194): 9–40.

Burniaux, J.-M., R. Duval, and F. Jaumotte (2004), "Coping with ageing: A dynamic approach to quantify the impact of alternative policy options on future labour supply in OECD countries," OECD Economics Department Working Paper No. 371, Paris.

Carmichael F. and S. Charles (1998), "The labour market costs of community care," *Journal of Health Economics* **17**: 747–65.

Carmichael F. and S. Charles (2003), "Benefit payments, informal care and female labour supply," *Applied Economics Letters* **10**(7): 411–15.

Carmichael F., S. Charles, and C. Hulme (2010), "Who will care? Employment participation and willingness to supply informal care," *Journal of Health Economics* **29**: 182–90.

Chang, C.F. and S.I. White-Means (1995), "Labour supply of informal caregivers," *International Review of Applied Economics* **9**(2): 192–205.

CISS (2005), "The challenges of aging and disability: Employment and insurance, international social security agreements," in The Americas Social Security Report 2006, Inter-American Conference on Social Security/CISS, Mexico City.

CISS (2008), "Evaluation of childcare and long-term care programs," in The Americas Social Security Report 2009, Inter-American Conference on Social Security/ CISS, 77–91.

Coe, N.B. and C.H. Van Houtven (2009), "Caring for mom and neglecting yourself? The health effects of caring for an elderly parent," *Health Economics* **18**(9): 991–1010.

CONEVAL (2011), Informe de la Evaluación Específica de Desempeño 2010-2011. Programa de Estancias infantiles para apoyar a madres trabajadoras. Mexico City, CONEVAL, Available online at http://www.coneval.gob.mx/Evaluacion/Eval_Especificas/Paginas/EED%202010-2011/EED-SEDESOL-2010-2011.aspx

Connelly, R. (1992), "The effect of child care costs on married women's labor force participation," *Review of Economics and Statistics* **74**: 83–90.

Connelly, R. and J. Kimmel (2003), "The effect of child care costs on the employment and welfare recipiency of single mothers," *Southern Economic Journal* **69**(3): 498–519.

Cox, D. and O. Stark (2005), "On the demand for grandchildren: Tied transfers and the demonstration effect," *Journal of Public Economics* **89**: 1665–97.

Ettner, S.L. (1996), "The opportunity costs of elder care," *Journal of Human Resources* **31**(1): 189–205.

Fevang, E., S. Kverndokk, and K. Røed (2008), "A model for supply of informal care to elderly parents," Health Economics Research Programme, University of Oslo, Working Paper 2008:12.

Fevang, E., S. Kverndokk, and K. Røed (2012), "Labor supply in the terminal stages of lone parents' lives," *Journal of Population Economics* **25**: 1399–1422.

Giménez-Nadal, J.I., M. Marcén, and J.A. Molina (2007), "How does the presence of children affect adult care and labour decisions? A psycho-economic approach," Institute for the Study of Labor, IZA Discussion Paper 2726.

Giménez-Nadal, J.I., M. Marcén, and R. Ortega (2010), "How do children affect parents' allocation of time?" *Applied Economics Letters* **17**(17): 1715–9.

Giménez-Nadal, J.I., M. Marcén, and R. Ortega (2012), "Substitution and presence effects of children on mothers' adult care time," *Journal of Family and Economic Issues* **33**: 2–10.

Hammer, L.B. and M.B. Neal (2008), "Working sandwiched-generation caregivers: Prevalence, characteristics, and outcomes," *The Psychologist-Manager Journal* **11**: 93–112.

Heitmueller, A. (2007), "The chicken or the egg? Endogeneity in labour market participation of informal carers in England," *Journal of Health Economics* 26: 536–59.

Juarez, L. (2009), "Crowding out of private support to the elderly: Evidence from a demogrant in Mexico," *Journal of Public Economics* **93**: 454–63.

Juarez, L. (2010), "The effect of an old-age demogrant on the labor supply and time use of the elderly and non-elderly in Mexico," *The B.E. Journal of Economic Analysis and Policy* **10**(1): 1–27.

Katz, S., T.D. Downs, H.R. Cash, and R.C. Grotz (1970), "Progress in development of the index of ADL," *The Gerontologist* **10**(1 Part 1): 20–30.

Leigh, A. (2010), "Informal care and labor market participation," *Labour Economics* **17**: 140–9.

Lilly, M.B., A. Laporte, and P.C. Coyte (2007), "Labor market work and home care's unpaid caregivers: A systematic review of labor force participation rates, predictors of labor market withdrawal, and hours of work," *The Milbank Quarterly* **85**(4): 641–90.

Meng, A. (2013), "Informal home care and labor-force participation of household members," *Empirical Economics* **44**(2): 959–79.

Michalopoulos, C. and Ph.K. Robins (2002), "Employment and child-care choices of single-parent families in Canada and the United States," *Journal of Population Economics* **15**: 465–93.

Miller, D.A. (1981), "The 'sandwich' generation: Adult children of the aging," *Social Work* **26**(5): 419–23.

Murrugarra, E. (2011), "Employability and productivity among older workers: A policy framework and evidence from Latin America," *Well-being and Social Policy* **7**(2): 53–99.

Nizalova, O. (2012), "The wage elasticity of informal care supply: Evidence from the health and retirement study," *Southern Economic Journal* **79**(2): 350–66.

Nocera, S. and P. Zweifel (1996), "Women's role in the provision of long-term care, financial incentives, and the future financing of long term care," in R. Eisen and F.A. Sloan, eds., *Long-term care: Economic issues and policy solutions*, Boston: Kluwer Academic Publishers.

OECD (2012), OECD Family Database, Paris: OECD (Available online at www.oecd.org/social/family/database).

Pommer, E., I. Woittiez, and J. Stevens (2007), "Comparing care. The care of the elderly in ten EU-countries," The Netherlands Institute for Social Research / SCP The Hague.

Puig, A., J.A. Pagán, and B.J. Soldo (2006), "Envejecimiento, salud y economía: La Encuesta Nacional sobre Salud y Envejecimiento en México," *El Trimestre Económico* **73**: 407–18.

Ribar, D.C. (1992), "Child care and the labor supply of married women," *Journal of Human Resources* **27**: 134–65.

Rubin, R.M. and S.I. White-Means (2009), "Informal caregiving: Dilemmas of sandwiched caregivers," *Journal of Family and Economic Issues* **30**: 252–67.

Schmitz, H. and M.A. Stroka (2013), "Health and the double burden of full-time work and informal care provision — Evidence from administrative data," *Labour Economics* **24**: 305–22.

Shelkey, M. and M. Wallace (2012), "Katz index of independence in activities of daily living (ADL)," in *Try this: Best practices in nursing care to older adults*, Issue 2, The Hartford Institute for Geriatric Nursing, New York University. Available online at http://consultgerirn.org/uploads/File/trythis/try_this_2.pdf

Spillman, B.C. and L.E. Pezzin (2000), "Potential and active family caregivers: changing networks and the 'sandwich generation,'" *The Milbank Quarterly* **78**(3): 347–74.

Tekin, E. (2007), "Child care subsidy, wages, and employment of single mothers," *Journal of Human Resources* **42**: 453–87.

van Gameren, E. (2008), "Labor force participation of the Mexican elderly: The importance of health," *Estudios Económicos* **23**(1): 89–127.

van Gameren, E. (2010), "Labor force participation by the elderly in Mexico," Centro de Estudios Económicos, El Colegio de México, Documento de Trabajo No. 2010-6.

Wolf, D.A. and B.J. Soldo (1994), "Married women's allocation of time to employment and care of elderly parents," *Journal of Human Resources* **29**(4): 1259–76.

Wong R., M. Espinoza, and A. Palloni (2007), "Adultos mayores mexicanos en contexto socioeconómico amplio: salud y envejecimiento," *Salud Pública de México* **49**: S436–S447.

Zúñiga Herrera, E. (2004), "Tendencias y características del envejecimiento demográfico en México," in Consejo Nacional de Población (CONAPO), *La situación demográfica de México*, 31–42.

Vol. 52 No. 2 (Nov, 2015), 149–183

Latin American Journal of Economics

THE EFFECTIVENESS OF PRENATAL CARE IN URUGUAY'S LOW-INCOME POPULATION: A PANEL DATA APPROACH*

Ana Inés Balsa**
Patricia Triunfo***

This paper studies the effectiveness of prenatal care on low-income women's birth outcomes. We analyze all births between 1995 and 2011 in Uruguay's largest public maternity ward. We use mother-specific first differences to circumvent biases due to time-invariant, unobserved heterogeneity and implement robustness checks that reduce concerns about time variant shocks and feedback effects. We find that adequate use of prenatal care, as defined by early initiation and at least 9 visits, decreases the probability of low birth weight by 6 percentage points and the probability of pre-term birth by 11 percentage points, and increases birth weight by 149 grams.

JEL classification: I12, J13, C14

Keywords: Prenatal care, low birth weight, low-income populations, first differences, panel data

1. Introduction

Preterm birth (< 37 weeks gestation) and low birth weight (< 2500 grams) are commonly used as proxies for infant health (McCormick, 1985; Institute of Medicine, 1985). Low birth weight (LBW) has been associated with increased morbidity and mortality during the life course, higher health care costs, lower educational attainment, and decreased lifetime income (Petrou *et al.*, 2000; Boardman *et al.*, 2002; Black *et al.*, 2007). Several authors have stressed that LBW serves as an important mechanism for the intergenerational transmission of economic status (Currie and Madrian, 1999; Grossman, 2000; Case *et al.*, 2004; Behrman and Rosenzweig, 2005; Currie and Moretti, 2005). To reduce the burden that LBW imposes on society, the medical community has underscored prenatal care as a key input. Through prenatal care, it is argued, mothers at risk of premature delivery or

* The authors would like to thank Graciela Sanromán and Partha Deb for their comments and methodological contributions and the anonymous LAJE referees who have improved this version of the paper. This work would not have been possible without data from the Perinatal Information System Division of the Epidemiology Area (National Information Unit of the Ministry of Public Health).

** University of Montevideo.

*** Faculty of Social Sciences, University of the Republic (Uruguay). Email: patricia.triunfo@cienciassociales.edu.uy.

doi 10.7764/LAJE.52.2.149

babies with intrauterine growth retardation (IUGR) can be identified, thus enabling a variety of medical, nutritional, and educational interventions aimed at reducing poor birth outcomes.

Estimation of the relationship between prenatal care and birth outcomes is complicated by difficulties in controlling for maternal characteristics associated with both the demand for prenatal care and the infant's health at birth. Without adequately controlling for health endowments, the mother's health habits, her propensity to engage in risky behaviors, or the extent to which the pregnancy is desired, an association between prenatal care and infant health cannot be regarded as causal. In the past 15 years the economic literature has proactively pursued the identification of such a causal relationship. Most investigations have exploited the association between exogenous variation in health care coverage and prenatal care use by employing either reduced-form models or two-stage least squares (2SLS) techniques (Kaestner, 1999; Brien and Swann, 2001; Currie and Grogger, 2002; Figlio *et al.*, 2009; Habibov and Fan, 2011). A few authors have used 2SLS with arguably exogenous instruments, such as input prices, the availability of prenatal clinics in the area (Grossman and Joyce, 1990; Gajate-Garrido, 2013), public transportation strikes in the county (Evans and Lien, 2005), and average distance to the nearest health facility (Awiti, 2014). Even after addressing confounders, this literature has been mixed regarding the effects of prenatal care on birth weight. The evidence is divided between those who find slight or no effects (Grossman and Joyce, 1990; Kaestner, 1999; Currie and Grogger, 2002; Kaestner and Lee, 2003; Evans and Lien, 2005; Figlio *et al.*, 2009) and those who find positive effects of significant magnitude (Rosenzweig and Schultz, 1983; Conway and Deb, 2005; Wehby *et al.*, 2009; Habibov and Fan, 2011; Awiti, 2014). This lack of consistency, together with other findings from clinical investigations (McDuffie *et al.*, 1996; Clement *et al.*, 1999; Villar *et al.*, 2001), has led some researchers to question whether the benefits of prenatal care have been "oversold" (Misra and Guyer, 1998).

A recent critique posits that past research has only been able to identify average effects, losing sight of the differential impact of prenatal care in different types of pregnancies (Conway and Deb, 2005). The results of randomized clinical trials, for example, are only valid externally for populations with low-risk pregnancies.[1] Moreover, 2SLS estimates local

1. Most randomized trials have been conducted in populations with low-risk pregnancies.

effects that are valid only for the population marginally affected by changes in the instrument. In addition, Conway and Deb argue that significant differences between healthy and risky pregnancies, when not explicitly accounted for, create a bimodal distribution of errors that downwardly biases the estimated effects of prenatal care.

Another problem with the literature is that, until recently, it almost exclusively focused on developed countries. The effectiveness of prenatal care may be quite different in developing countries, where women are, in general, less informed about the health consequences of certain conditions and behaviors, and have, arguably, fewer resources to address nutritional and hygiene needs. The role of the health care provider, particularly in low-income contexts, may be critical to promote healthy pregnancies and decrease the incidence of LBW. In line with this argument, preliminary evidence for Uruguay shows that the transition from no prenatal care to nine prenatal visits significantly increases the birth weight of the child (Jewell and Triunfo, 2006; Jewell *et al.*, 2007). More recent research based in Argentina, Azerbaijan, Cebu, and Kenya provides evidence of the strong effects of prenatal care (Wehby *et al.*, 2009; Habibov and Fan, 2011; Gajate-Garrido 2011; Awiti, 2014).

This paper analyzes the impact of prenatal care on the likelihood of preterm birth and LBW in a low-income population in Uruguay, a middle-income country in South America with a population of 3.3 million. Our estimation strategy addresses time-invariant, unobserved confounders by exploiting intra-mother variation in prenatal care for women who had at least two births between 1995 and 2011 in the main maternity hospital in the capital city Montevideo (Pereira Rossell Hospital). We also perform several robustness checks that reduce concerns about biases due to time-varying, unobserved heterogeneity and to feedback effects from previous pregnancies.

Our analysis sheds new light on the effectiveness of prenatal care for poor populations in developing countries. The evidence is very timely for Uruguay, given the recent efforts by the Ministry of Public Health (MPH) to improve the coverage of prenatal care in the country. Since 2010, the Uruguayan government has been financially rewarding providers affiliated with the National Social Health Insurance system for increasing the fraction of pregnant mothers who initiate care during the first trimester and attend at least six prenatal visits. While all health-care providers are, in principle, eligible to receive supplemental payments, in general public providers are not receiving the incentive payments because most of the population they serve is not affiliated with the National Social Health Insurance system.[2] Our results have clear public policy implications. If the State Health Services Administration in Uruguay, which offers health services to low SES women, were to achieve the prenatal care goals proposed by the MPH, the rate of low birth weight in this population would drop by 4 percentage points, from 10% to 6.2%, increasing the number of newborns with adequate birth weight by 800 per year.[3] The effects would be even stronger (two additional percentage points) if the goals were aimed at achieving nine rather than six prenatal visits in a full-term pregnancy. The number of children born with adequate weight would increase, in that case, by 1,200 per year.

2. Conceptual Framework and Methodology

The production of birth weight is usually modeled in economics as the result of a process of parental utility maximization. Parents' utility is a function of their children's well-being, which depends directly on the children's health. Conditional on genetic endowments and household resources, parents are indirect producers of children's health and decide which inputs to invest in to maximize their children's health status. These inputs include the use of prenatal care, use of substances during pregnancy, exercise, and nutrition, among others (Grossman, 2000). As mentioned earlier, the analyst's inability to observe the full set of maternal preferences, resources, and information involved in the household production of health may lead to estimation biases. Unobserved characteristics such as the health endowment of the fetus and/or the mother, the mother's health habits, her propensity to engage in risky behavior, or the extent to which the pregnancy is desired, could bias the estimates if their influence is not taken into consideration.

In this paper, we address this endogeneity using differencing techniques. Our methodology exploits the availability of longitudinal information

2. Uruguay has a mixed health insurance system. The population formally engaged in the labor market and their families are covered by a National Social Health Insurance that provides services either through private or public providers. Most beneficiaries are enrolled with private providers. The public provider, the State Health Services Administration (ASSE), covers low-income populations that are not formally inserted in the labor market. This population shows the lowest rates of initiation of care during the first trimester and the highest non-compliance with recommended standards of care.

3. This back-of-the-envelope calculation is based on the approximately 20,000 births in the public health care sector in Uruguay per year.

for the same mother over several pregnancies. The underlying model is of the form:

$$Y_{ij} = a_0 + a_1 CP_{ij} + X'_{ij}\beta + a_i + \varepsilon_{ij} \tag{1}$$

where Y_{ij} reflects the outcome of mother i's pregnancy j (preterm birth, birth weight, very low birth weight, and low birth weight), j denotes the birth order, CP_{ij} is an indicator of the adequacy of prenatal care, and X_{ij} includes other determinants of the newborn's health (such as mother's age, education, marital status, tobacco use, body mass index, history of past births, and quarter and year of pregnancy). The term α_i captures the time-invariant unobserved heterogeneity in i, namely, personality characteristics of the mother that affect her habits, her involvement in risky behavior, her health endowment, her knowledge about the benefits of prenatal care, her preferences, and so forth. Finally, ε_{ij} is an idiosyncratic error term independent of α_i and other explanatory variables.

A naive estimate of a_1 resulting from the regression of the outcome variable Y_{ij} on the indicator of adequacy of prenatal care CP_{ij} is potentially biased (even after adjusting for other controls X_{ij}) if it does not account for the unobserved heterogeneity component, α_i, which is associated both with the explanatory variable of interest and with the dependent variable. This has been a common estimation error in the biomedical literature, which has resulted in unreliable estimates of the effectiveness of prenatal care.

We begin our analysis by projecting deviations in neonatal outcomes on deviations in prenatal care that happen across the same mother's different pregnancies (adjusting for within variations in other characteristics). In order to eliminate the potential correlation between the mother-specific fixed effect α_i and inputs in the production function of child health, the methodology requires a transformation of the data into within-mother deviations. We work with first differences in our core model[4], and for sensitivity, we also re-estimate the models using the "within transformation" (fixed effects) and orthogonal deviations

4. This transformation involves subtracting the observation in $j-1$ from that in j for the same mother.

(Arellano and Bover, 1995). Once the data are transformed, the model identifies the effect of interest, a_1, by getting rid of the idiosyncratic time invariant term α_i:

$$Y_{ij} - Y_{ij-1} = a_1(CP_{ij} - CP_{ij-1}) + a_2(X_{ij} - X_{ij-1}) + \varepsilon_{ij} - \varepsilon_{ij-1} \qquad (2)$$

The nice feature of this method relative to 2SLS is that it estimates average treatment effects. One of the problems with the extant literature on the effectiveness of prenatal care is that it tends to rely exclusively on the population of compliers, i.e., those who increase their use of prenatal care when confronted with an exogenous policy shock (an increase in health care coverage, for example). These local treatment effects may provide a distorted picture of the effectiveness of prenatal care if the impacts are heterogeneous across the different subpopulations. If policy changes do not modify, in the margin, the behavior of those most likely to benefit from prenatal care, the 2SLS estimates will underestimate its impact. Moreover, even if the group of compliers includes complicated and normal pregnancies, combining them in a single 2SLS estimation may yield bimodal residuals that will result in insignificant estimates. Using a finite mixture model, Conway and Deb (2005) find estimates of prenatal care that have a consistent, substantial effect on normal pregnancies. Using a Monte Carlo experiment, they show that ignoring even a small proportion of complicated pregnancies can cause prenatal care to appear as insignificant.

While fixing biases due to time-invariant, unobserved heterogeneity, the differencing technique proposed in (2) may fail to produce consistent estimates in two scenarios: a) if there are unobserved time-variant shocks associated with both the use of prenatal inputs and birth outcomes (a problem we refer to as time-variant, unobserved heterogeneity); and b) if past birth outcomes affect the current demand for prenatal inputs (a problem we refer to as feedback effects). The former would include any changes in preferences, resources, or information between deliveries that are not captured by the time-variant adjustors used in the analysis. For example, this would be the case if an unobserved negative shock on the fetus' health endowment decreased the expected prenatal outcomes and led the mother to increase the use of prenatal care. Or, if the government implemented an information campaign that encouraged the use of prenatal care as well as other changes in maternal behavior. The second problem would occur if past shocks affected the contemporaneous demand for inputs. For example, a mother may react to an adverse shock to a previous pregnancy (a pregnancy that ended in a preterm birth or that had some risk of miscarriage) by increasing the demand for inputs or for the quality of inputs in the current pregnancy. In either case, working with deviations will not lead to consistent estimates of the coefficients of interest.

To address these problems Arellano and Bond (1991) devised a difference-GMM technique that uses the level of the explanatory variable lagged two periods as the instrument for the first difference.[5] This approach is, in principle, able to address feedback effects and endogeneity due to time-variant, unobserved heterogeneity. We initially attempted to take this avenue, but as in Abrevaya (2006), the two-period-lag instruments were too imprecise. Thus, in this paper we proceed to formally address the feedback effects problem, which requires using one-period lags of the explanatory variables as instruments, and discuss later why we think the problem of time-variant, unobserved heterogeneity may be, at most, moderate.

To see the feedback effects problem formally, suppose that the model is as in (1), but past shocks pre-determine the level of inputs in t:

$$E(CP_{ij}, \varepsilon_{ij-1}) \neq 0 \tag{3}$$

Under this assumption, the first difference transformation in (2) generates an endogenous relationship between the deviations in prenatal care and the differenced error term. We propose to address this feedback problem by running GMM on first differences (difference-GMM) and by using a one-period (and eventually deeper) lag(s) of the predetermined variable as "GMM-style" instrument(s) of the contemporary deviations in that variable (Holtz-Eakin *et al.*, 1988; Arellano and Bond, 1991; Roodman, 2009). Specifically, we use the level of prenatal care in pregnancy $j-1$, and deeper lags when available, as instruments for the difference in prenatal care use between pregnancies j and $j-1$. The

5. Abrevaya (2006) recognized these problems and attempted a similar correction while analyzing the effect of tobacco use on birth outcomes.

orthogonality conditions in our GMM model are: $E[X_{i1}\ (\varepsilon_{i2} - \varepsilon_{i1})] = 0$ for mothers with two deliveries in the period, $E[X_{i1}\ (\varepsilon_{i2} - \varepsilon_{i1})] = E[X_{i1}\ (\varepsilon_{i3} - \varepsilon_{i2})] = E[X_{i2}\ (\varepsilon_{i3} - \varepsilon_{i2})] = 0$ for mothers with three deliveries, and so forth with mothers with more than three deliveries.

Our first specification assumes no feedback effects: it relies on the assumption that past shocks are orthogonal to the current demand for prenatal inputs. Next, we allow for feedback effects to play a role and estimate the model using the lagged levels of prenatal care as instruments of the first difference in prenatal care. All the regressions control for the year of birth dummies and compute robust standard errors that are clustered at the mother's level. We run two specifications of the model: one without adjusting for the duration of the pregnancy, and the other controlling for the number of weeks of gestation at delivery. This latter specification accounts for the fact that the effects of prenatal care on birth weight can occur through the probability of reaching full term. In the GMM specification, we instrument deviations in gestational weeks with two lags of the number of gestational weeks in levels.

3. Data

We analyze births registered in the Perinatal Information System (CLAP-OP-OMS, 1999) of the Pereira Rossell Hospital between 1995 and 2011. Pereira Rossell is the public teaching hospital for the only public university in the country, and is part of the State Health Services Administration (ASSE). The hospital is a referral center for acute care of mothers and children for the whole country, concentrating 70% of the births that take place in public wards in Montevideo, 33% of all births in Montevideo, and 15% of births nationwide. The hospital serves women who have no access to private insurance through their employment or who cannot afford individual insurance. These women are entitled to access to prenatal and obstetric care free-of-charge at public facilities.

The Perinatal Information System at the Pereira Rossell Hospital covers approximately 98.5% of all births that take place in the hospital, according to national birth registries. These data are quite unique in that they allow for identification of mothers over a period of 17 years. The dataset is larger than similar ones used in other medical and epidemiological studies, and provides information on a population of women who have not been studied intensively, with SES, cultural, and geographic differences relative to women in developed countries.

The outcomes in our analysis are preterm birth, birth weight (BW), low birth weight (LBW), and very low birth weight (VLBW). We consider a delivery to be preterm if it occurs before the 37th week of gestation. BW is a continuous variable in grams. LBW is a binary variable that takes the value of 1 if the birth weight is less than 2,500 grams and 0 otherwise. VLBW is a binary variable that takes the value of 1 if the birth weight is less than 1,500 grams and 0 otherwise.

We analyze three alternative measures of prenatal care. Our core measure is based on the Kessner Index, a widely used indicator of adequacy of care (Kotelchuck, 1994). According to Kessner's criterion, a woman has "adequate" prenatal care if she has her first visit during the first trimester (week 13 or earlier) and has at least nine visits at term, or between 4 and 8 visits in the case of a preterm birth. Prenatal care is deemed "inadequate" if the mother initiates the control visits in the third trimester or if care was initiated before but she has fewer than 4 control visits by the time of delivery, or between 1 and 3 visits when the birth is premature. All other combinations of initiation and visits belong to an "intermediate" category.

The second categorization of prenatal care is based on MPH guidelines established in 2010, which serve as the reference for incentive payments received by providers. The variable takes the value of 1 if the woman initiated prenatal care in the first trimester of her pregnancy and attended at least 6 control visits by the time of delivery. Finally, we analyze the timing of initiation of prenatal care, another variable analyzed in the literature that takes the value of 1 if visits were initiated in the first trimester and 0 otherwise.

We use the following covariates to adjust for mother or pregnancy-specific observed heterogeneity. Tobacco use is an indicator of any smoking, as reported by the mother at her first prenatal visit. Smoking during pregnancy has been associated with low birth weight (Permutt and Hebel, 1989; Veloso da Veiga and Wilder, 2008; Reichman *et al.*, 2009). Previous research has also shown an inverse, U-shaped relationship between maternal age and birth weight: Women who are either under the age of 26 or above 35 have higher rates of low birth weight children compared to other childbearing women (Abel *et al.*, 2002). To capture this pattern, we include five categories of maternal age: under 17, between 17 and 19, between 20 and 34 (optimal age group), between 35 and 39, and over 40. As proxies for socioeconomic status, we consider mothers' marital status (married, single, cohabitating, other), and education (whether the mother completed primary education, middle school, or high school). Among the mother's epidemiologic risk factors, we consider body mass index (BMI) before the pregnancy (underweight, overweight, or obese) and the presence of hypertension, preeclampsia, and eclampsia in each pregnancy. Both high BMI and chronic hypertension have been associated with low birth weight (Ehrenberg *et al.*, 2003, Haelterman *et al.*, 1997). Finally, the epidemiology literature shows that the experience of previous births is associated with anatomical changes that may impact the health of a newborn (Khong *et al.*, 2003). Among these variables, we consider the number of prior births, episodes of mortality in prior deliveries, and prior abortions. For biological reasons, girls generally weigh less than boys (Thomas *et al.*, 2000), so we include a dummy variable that equals one if the newborn is a boy.

Because of the differencing methodology used in this paper, we work with a sample of low-SES Uruguayan women who gave birth to at least two children between 1995 and 2011. The data include information about the mother, the pregnancy, and the newborn's health. Of the 143,228 total births registered in the hospital in the period (about 8,400 births a year), 995 were discarded because of unviable pregnancies (less than 25 weeks of gestation or birth weights below 500 grams), 3,343 were not considered because of multiple pregnancies, and 7,227 were not included because proper identification of the mother was lacking. Of the remaining observations (131,663), we discarded those with inconsistent information or missing values for relevant variables and considered only births to mothers who delivered at least twice during the period (33% of births). Finally, we excluded pregnancies that did not show variation in the use of prenatal care across the same mother.

The final sample comprises 40,729 births. Altogether, there are 17,278 mothers in the sample, of which 12,719 had two births, 3,335 had three, 911 had four, 253 had five, and 60 had six or more. Table 1 presents descriptive statistics for mothers with at least two deliveries in the period of analysis. The first and second columns show means and standard deviations for first pregnancy characteristics, while the third and fourth present data on all pregnancies. The rate of prematurity in all deliveries is 14%, the average birth weight is 3,145 grams, 1.5% of the newborns have very low birth weight, and 10% have low birth weight. Only 13% of all pregnancies have a proper

Table 1. Summary statistics, analysis sample 1995-2011

Mothers with at least two deliveries

Variable	All pregnancies	
	Mean	Standard deviation
Preterm birth	0.139	0.346
Birth weight	3,145	567
Very low birth weight (VLBW)	0.015	0.122
Low birth weight (LBW)	0.100	0.300
Mother had adequate prenatal care (MPH)	0.221	0.415
Initiated prenatal care in first trimester	0.264	0.441
At least 6 prenatal care visits	0.496	0.500
Inadequate prenatal care (Kessner)	0.436	0.496
Intermediate prenatal care (Kessner)	0.430	0.495
Adequate prenatal care (Kessner)	0.134	0.341
Age < 17	0.046	0.210
Age 17-19	0.163	0.369
Age 20-34	0.705	0.456
Age 35-39	0.068	0.251
Age > 39	0.019	0.135
Marital status: cohabitation	0.183	0.387
Marital status: single	0.597	0.490
Marital status: other	0.201	0.401
Marital status: married	0.018	0.133
Education: did not finish primary school	0.155	0.362
Education: completed primary school	0.625	0.484
Education: completed middle school	0.191	0.393
Education: completed high school	0.029	0.167
Smoker	0.395	0.489
Smoking status missing	0.000	0.000
Mother is underweight (BMI < 18.5)	0.060	0.238
Mother is overweight ($25 \leq$ BMI < 30)	0.078	0.269
Mother is obese (BMI ≥ 30)	0.035	0.183
Missing BMI	0.494	0.500
Previous stillbirth	0.036	0.217
Previous abortions	0.244	0.644
Parity (max $= 16$)	2.335	2.152
First child	0.184	0.387
Hypertension	0.019	0.138
Hypertension missing	0.087	0.281
Preeclampsia	0.018	0.133

Table 1. (continued)

Variable	All pregnancies	
	Mean	Standard deviation
Preeclampsia missing	0.087	0.282
Eclampsia	0.001	0.033
Eclampsia missing	0.088	0.283
Child's gender: Male	0.512	0.500
Trimester of birth: 1	0.243	0.429
Trimester of birth: 2	0.251	0.434
Trimester of birth: 3	0.264	0.441
N	40,729	

Source: Authors' estimations. Data from the Perinatal Information System of Pereira Rossell Hospital (Ministry of Public Health).

follow-up according to the Kessner criterion, 43.6% are inadequately followed-up, and the rest are in between. Barely 22% of pregnancies meet the prenatal care goals set by the Uruguayan Ministry of Public Health. While half of the women attend the target six control visits by the time of delivery, only 26% initiate care in the first trimester. These figures are quite surprising when considering that pregnancy care is free in Uruguay and that there are few geographic barriers to obstetric care facilities. Similar behavior, however, has been found among persons eligible for public assistance programs in the United States (Currie and Grogger, 2002; Kaestner and Lee, 2003). Most women in the sample are between 20 and 34 years old, unmarried, and have not finished middle school. One out of three mothers smokes during the pregnancy.

In Table A1 of the appendix we compare the first pregnancy characteristics of mothers in the analysis sample with those of the excluded mothers (i.e., mothers observed only once during the period). Mothers in the analysis sample are younger, less educated, more likely to be smokers, and less likely to have adequate prenatal care during their first observed pregnancy. On the other hand, they are less likely to have low weight or very low weight children, are less likely to be overweight or obese, and have a lower likelihood of having hypertension and preeclampsia during the pregnancy. While mothers with two or more deliveries are clearly different than mothers with only one delivery, the above comparison does not suggest obvious discrepancies in the quality of pregnancies or in mothers' social vulnerability between the two samples. For external validity purposes, we also compare women delivering at Pereira Rossell with low-income women delivering at other public hospitals in Uruguay. We make this comparison for the 2000-2011 period, as we do not have good information on other hospitals before the year 2000.[6] While the age distribution is similar in both samples, women delivering at Pereira Rossell Hospital are slightly less educated and less likely to be married. They are also more likely to have premature or low birth weight children. This comparison suggests that women in our study sample may be somehow more vulnerable than other low-income women in Uruguay and that our results are more likely to apply to the poorest fractions of the population.

4. Results

4.1 Kessner criterion for impact of adequate prenatal care

Table 2 shows the results of the estimation when the adequacy of prenatal care is defined in terms of the Kessner index. Each column in Table 2 depicts the estimates of a linear regression model on a panel of observations that have been transformed to within-mother first differences. The defined categories of prenatal care already acknowledge that reduced duration of gestation truncates the time available to make visits, thus avoiding a problem of mechanical reverse causation.[7] Results show large effects of prenatal care on all birth outcomes considered. An adequate use of prenatal care decreases the likelihood of preterm birth by 11.3 percentage points, the likelihood of VLBW by 1.1 percentage points, and LBW by 6 percentage points, and increases birth weight by 149 grams. Even if prenatal care is not fully adequate in the sense of Kessner, women initiating care before the third trimester and showing at least four prenatal visits by the end of the pregnancy (intermediate prenatal care) are 7.6 percentage points less likely to experience a preterm birth and 3.4 percentage points less likely to deliver a baby below 2,500 grams than women with inadequate use. Once we adjust for gestational age (columns 3, 5, and 7), the estimated impact of adequate and intermediate care on LBW and birth weight decreases

6. Results are available upon request.

7. Women who deliver before the 37th week of gestation are categorized as compliers with adequate standards of care if they have completed the number of visits that corresponds to that gestation week.

Table 2. Effects of adequate prenatal care (Kessner index) on birth outcomes

Estimation using mother-specific first differences ($N = 23{,}451$)

	Preterm birth	Birth weight		VLBW		LBW	
		No control for gestational age	Adjusting for gestational age	No control for gestational age	Adjusting for gestational age	No control for gestational age	Adjusting for gestational age
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intermediate prenatal care	-0.076***	76.341***	28.952***	-0.003	0.008***	-0.034***	-0.009**
	(0.006)	(7.274)	(6.154)	(0.002)	(0.002)	(0.005)	(0.004)
Adequate prenatal care	-0.113***	148.627***	67.858***	-0.011***	0.007***	-0.060***	-0.019***
	(0.008)	(11.098)	(9.210)	(0.003)	(0.003)	(0.006)	(0.006)
Gestational age (weeks)			104.120***		-0.024***		-0.053***
			(1.729)		(0.001)		(0.001)
Smoker	-0.001	-42.311***	-37.019***	0.003	0.002	0.016***	0.013***
	(0.007)	(9.113)	(7.582)	(0.003)	(0.002)	(0.006)	(0.005)
Age < 17	0.038**	-48.414**	-9.845	0.001	-0.008	0.022*	0.002
	(0.015)	(20.006)	(16.495)	(0.006)	(0.005)	(0.013)	(0.011)
Age 17-19	0.021***	-47.597***	-23.942***	0.001	-0.005*	0.015**	0.003
	(0.008)	(10.738)	(9.005)	(0.003)	(0.003)	(0.007)	(0.006)
Age 35-39	0.020	-3.309	10.875	-0.003	-0.006	0.006	-0.002
	(0.012)	(17.141)	(14.652)	(0.005)	(0.004)	(0.010)	(0.009)
Age > 40	0.044**	-34.030	-8.404	0.000	-0.006	0.009	-0.005
	(0.022)	(32.483)	(28.320)	(0.009)	(0.008)	(0.018)	(0.017)
Cohabitation	-0.000	-24.854**	-22.235**	0.005	0.005	0.010	0.009
	(0.009)	(12.289)	(10.225)	(0.003)	(0.003)	(0.008)	(0.007)
Single	0.008	-33.558**	-30.688***	0.006	0.005	0.008	0.006
	(0.011)	(14.135)	(11.626)	(0.004)	(0.004)	(0.009)	(0.008)
Other marital status	-0.003	-10.512	-22.318	-0.007	-0.004	0.025	0.031**
	(0.019)	(25.543)	(21.784)	(0.007)	(0.006)	(0.016)	(0.014)

Table 2. (continued)

	Preterm birth	Birth weight		VLBW		LBW	
		No control for gestational age	Adjusting for gestational age	No control for gestational age	Adjusting for gestational age	No control for gestational age	Adjusting for gestational age
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Finished primary	0.004	9.741	14.374	0.003	0.002	-0.002	-0.005
	(0.011)	(13.780)	(12.111)	(0.004)	(0.004)	(0.009)	(0.008)
Finished middle school	-0.011	23.473	15.684	-0.002	0.000	-0.013	-0.009
	(0.013)	(17.305)	(15.215)	(0.005)	(0.005)	(0.011)	(0.010)
Finished high school	-0.014	5.108	-2.703	-0.000	0.001	0.008	0.012
	(0.019)	(27.482)	(23.003)	(0.008)	(0.007)	(0.016)	(0.014)
Mother is underweight	-0.008	-15.606	-23.247**	0.003	0.005	-0.001	0.003
	(0.012)	(14.180)	(11.718)	(0.004)	(0.004)	(0.010)	(0.009)
Mother is overweight	0.018**	8.662	20.315*	0.002	-0.000	-0.001	-0.007
	(0.009)	(12.907)	(11.261)	(0.003)	(0.003)	(0.007)	(0.006)
Mother is obese	0.036***	18.993	46.074**	0.000	-0.006	0.026**	0.012
	(0.013)	(21.721)	(18.530)	(0.005)	(0.005)	(0.010)	(0.009)
BMI missing	0.031***	-30.879***	-6.060	0.007***	0.001	0.019***	0.007*
	(0.005)	(7.193)	(6.036)	(0.002)	(0.002)	(0.004)	(0.004)
Previous stillbirth	-0.068***	134.697***	61.831***	-0.050***	-0.033***	-0.069***	-0.032**
	(0.018)	(29.370)	(21.011)	(0.011)	(0.009)	(0.018)	(0.014)
N previous abortions	-0.003	-5.511	-6.697	0.001	0.001	0.000	0.001
	(0.005)	(8.694)	(7.193)	(0.003)	(0.002)	(0.005)	(0.004)
Parity	0.003	8.765	13.657***	0.001	0.000	-0.004	-0.007**
	(0.004)	(5.547)	(4.682)	(0.002)	(0.001)	(0.003)	(0.003)
First pregnancy	0.009	-94.384***	-91.672***	0.008***	0.008***	0.035***	0.033***
	(0.007)	(9.870)	(8.085)	(0.003)	(0.003)	(0.006)	(0.006)
Hypertension	0.004	-7.715	0.738	-0.001	-0.003	0.018	0.014

Table 2. (continued)

	Preterm birth	Birth weight		VLBW		LBW	
		No control for gestational age	Adjusting for gestational age	No control for gestational age	Adjusting for gestational age	No control for gestational age	Adjusting for gestational age
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	(0.018)	(28.988)	(24.927)	(0.008)	(0.007)	(0.016)	(0.014)
Hypertension missing	0.040	24.955	8.005	-0.015	-0.011	0.026	0.035
	(0.064)	(95.200)	(72.608)	(0.026)	(0.023)	(0.063)	(0.055)
Preeclampsia	0.148***	-215.457***	-113.140***	0.033***	0.010	0.110***	0.057***
	(0.021)	(32.217)	(26.078)	(0.011)	(0.010)	(0.020)	(0.017)
Preeclampsia missing	0.048	-51.377	-38.274	-0.037	-0.040	-0.006	-0.013
	(0.083)	(124.258)	(88.697)	(0.045)	(0.038)	(0.086)	(0.075)
Eclampsia	0.194**	-304.866***	-117.552	0.133**	0.090**	0.178**	0.082
	(0.077)	(114.788)	(75.523)	(0.052)	(0.044)	(0.083)	(0.069)
Eclampsia missing	-0.117	23.606	11.159	0.052	0.055	-0.025	-0.018
	(0.091)	(140.376)	(96.250)	(0.049)	(0.043)	(0.098)	(0.084)
Male	0.006	115.458***	117.403***	-0.001	-0.002	-0.014***	-0.015***
	(0.004)	(5.744)	(4.926)	(0.002)	(0.001)	(0.004)	(0.003)
Trimester of birth:1	0.008	14.210	24.125***	-0.006***	-0.008***	-0.006	-0.011**
	(0.006)	(8.649)	(7.441)	(0.002)	(0.002)	(0.005)	(0.005)
Trimester of birth:2	0.014**	7.446	20.333***	0.001	-0.002	-0.005	-0.011**
	(0.006)	(8.331)	(7.100)	(0.002)	(0.002)	(0.005)	(0.005)
Trimester of birth:3	0.001	7.992	11.347	-0.000	-0.001	-0.011**	-0.012***
Year fixed effects	yes	yes	yes	yes	yes	yes	yes

Source: Authors' estimations. Data from the Perinatal Information System of Pereira Rossell Hospital (Ministry of Public Health).
Note: Clustered standard errors in parentheses. * p < 0.1, ** p < .05, *** p < .01.

substantially, suggesting that prenatal care improves birth weight mainly through a reduction in the likelihood of a preterm birth. Our estimates show that 4 percentage points out of the 6-percentage-point decrease in uncontrolled LBW are due to the fall in prematurity.

Both the probability of LBW and the likelihood of preterm birth are higher for teenage mothers (19 years or younger), for mothers with prior obesity, and for mothers with preeclampsia or eclampsia during pregnancy; they are lower for mothers who had a still-birth in a prior pregnancy. LBW is also positively associated with being a smoker, with a first pregnancy, and with a female baby.

We repeat the analysis using within-mother fixed effects and orthogonal differences instead of first differences, and the results are strongly robust to these variations.

4.2 Other measures of prenatal care adequacy: early initiation and MPHC guidelines

Tables 3 and 4 present the findings for alternative measures of adequate prenatal care. Table 3 shows the effects of initiation of care during the first trimester on the likelihood of preterm delivery and LBW. When compared to the results in Table 2, these estimates give a sense of the relative importance of early initiation versus the number of control visits in the overall effect of adequacy of care. Early initiation has some impact on the likelihood that the pregnancy reaches full term, but this impact is much smaller than the aggregate effect of early initiation plus an adequate number of visits: initiation during the first trimester decreases a preterm birth by 1.3 percentage points. This finding suggests that the positive effect of prenatal care on the duration of the pregnancy and on birth weight would be due mostly to an adequate number of visits. It also highlights the importance of quantifying the number of visits in addition to the time of initiation in studies analyzing the effectiveness of prenatal care.

Table 4 shows the effects of compliance with the standards of prenatal care set by the Uruguayan Ministry of Public Health. Complying with the MPH standards decreases the likelihood of prematurity by 4.8 percentage points. The effects on LBW are also statistically significant and large. Initiating care during the first trimester and making at least six visits by the end of the pregnancy reduces the likelihood of LBW by 4 percentage points (a 40% decrease).

Table 3. Effects of early initiation of prenatal care on birth outcomes

Estimation using mother-specific first differences

	Preterm birth	Birth weight		VLBW		LBW	
		No control for gestational age	Adjusting for gestational age	No control for gestational age	Adjusting for gestational age	No control for gestational age	Adjusting for gestational age
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Initiation during 1st trimester	-0.013**	31.997***	22.650***	-0.005**	-0.003*	-0.017***	-0.013***
	(0.006)	(7.646)	(6.216)	(0.002)	(0.002)	(0.005)	(0.004)
Gestational weeks		104.858***		-0.023***		-0.054***	
		(1.718)		(0.001)		(0.001)	

Source: Authors' estimations. Data from the Perinatal Information System of Pereira Rossell Hospital (Ministry of Public Health).
Note: All estimations include the full set of controls depicted in Table 2. Clustered standard errors in parentheses. * $p < 0.1$, ** $p < .05$, *** $p < .01$.

Table 4. Effects of adequacy of prenatal care according to Uruguayan MPH guidelines

Estimation using mother-specific first differences

	Preterm birth	Birth weight		VLBW		LBW	
		No control for gestational age	Adjusting for gestational age	No control for gestational age	Adjusting for gestational age	No control for gestational age	Adjusting for gestational age
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Adequate prenatal care (MPH)	-0.048***	80.365***	36.746***	-0.019***	-0.009***	-0.040***	-0.018***
	(0.006)	(8.126)	(6.640)	(0.002)	(0.002)	(0.005)	(0.004)
Gestational weeks		104.510***		-0.023***		-0.053***	
		(1.714)		(0.001)		(0.001)	

Source: Authors' estimations. Data from the Perinatal Information System of Pereira Rossell Hospital (Ministry of Public Health).
Note: All estimations include the full set of controls depicted in Table 2. Clustered standard errors in parentheses. * $p < 0.1$, ** $p < .05$, *** $p < .01$.

The comparison between the Kessner effects and those estimated with the MPH guidelines suggests that increasing the target number of control visits beyond those required by the MPH may lead to more pronounced decreases in the probability of LBW. Note that there is about a 2 percentage-point-difference between the impact of the Uruguayan MPH guidelines on LBW (-0.04) and the coefficient on the Kessner measure of adequate care (-0.06). The additional three visits required in the Kessner measure of adequacy appear to have a large impact on birth outcomes. Moreover, much of the beneficial impact of this increased number of control visits seems to operate through a smaller probability of preterm birth.

5. Robustness Analysis

The main contribution of first differencing techniques is to reduce biases due to time-invariant, unobservable heterogeneity. However, as mentioned before, this technique may fail to produce consistent estimates in two scenarios: a) if there are feedback effects in the decision to use prenatal care; and b) if there are time-variant shocks associated with both the use of prenatal inputs and birth outcomes. As we show below, several checks support the robustness of our core findings even when considering these other confounders.

5.1 Feedback effects in the demand for prenatal care

Results for the feedback effects model are presented in Table 5. The likelihood of a preterm birth in this specification falls by 12.6 percentage points when prenatal care is adequate and by 7.9 percentage points when care is intermediate, relative to inadequate care. These coefficients are very similar to those in the model in first differences, suggesting a small role for feedback effects. Regarding the effects on LBW, the coefficient on adequate care is now -0.058 and that on intermediate care is -0.029. While slightly smaller than the coefficients in the core model, the estimates are still significant, large, and not statistically distinguishable from the magnitudes previously described.

Table 5. Effects of adequacy of prenatal care, as defined by the Kessner index, on birth outcomes

Difference GMM with feedback effects

	Preterm birth	Birth weight		VLBW		LBW	
		No control for gestational age	Adjusting for gestational age	No control for gestational age	Adjusting for gestational age	No control for gestational age	Adjusting for gestational age
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intermediate prenatal care (Kessner)	-0.079***	47.857***	-9.380	0.001	0.009**	-0.029***	0.002
	(0.009)	(11.805)	(13.099)	(0.003)	(0.005)	(0.007)	(0.009)
Adequate prenatal care (Kessner)	-0.126***	120.947***	21.275	-0.004	0.010	-0.058***	-0.004
	(0.012)	(17.735)	(20.575)	(0.005)	(0.008)	(0.010)	(0.014)

Source: Authors' estimations. Data from the Perinatal Information System of Pereira Rossell Hospital (Ministry of Public Health).
Note: All estimations include the full set of controls depicted in Table 2. Clustered standard errors in parentheses. * $p < 0.1$, ** $p < .05$, *** $p < .01$.

5.2 Time-variant heterogeneity

Time-variant shocks would include any changes in preferences, endowments, or information between deliveries that are not captured by the time-variant adjustors included in the analysis. Our models already address time-variant heterogeneity stemming from aggregate shocks and from mother's health endowments. We capture time-variant aggregate shocks, such as information campaigns, improvements in the quality of prenatal care, or a better economy by introducing year fixed effects.[8]

In our core specification we account for the mother's pregnancy-specific health endowments by including controls for preeclampsia, eclampsia, and hypertension, the three most important pregnancy-related conditions. To assess whether the estimates are sensitive to maternal health shocks, we rerun the above models excluding these conditions. We find that prenatal care coefficients are not affected by these exclusions (results are available upon request). Overall, we are not particularly concerned about biases due to mother-specific or fetus-specific adverse shocks, because we expect them to shift our estimates towards zero. A less healthy mother or a less healthy fetus would be associated with worse pregnancy outcomes and with more use of prenatal care. In such a case, our findings would be conservative.

We are more concerned about time-variant unobserved heterogeneity not uniform across all women that may potentially inflate our estimates. For example, both the demand for health care and fetus health may be compromised by a negative income shock at the household level. While prenatal care is provided free of charge in Uruguay, there are other costs involved (i.e., transportation, child care) that could decrease its use. Alternatively, women who engage in alcohol or drug consumption between two pregnancies are expected to have poorer birth outcomes and, at the same time, to decrease their demand for prenatal care (due to indolence or fear of such consumption being detected).

While we are not able to address each of these conjectures directly[9], we check for robustness by rerunning the analysis for different periods of time and different samples of women. Our assumption is that the

8. This approach is valid only if the time-variant effects are constant across all women in the population.

9. Abrevaya (2006) recognized these problems while analyzing the effect of tobacco on birth outcomes, and suggested using a difference-GMM model with the level of the explanatory variable lagged two periods as the instrument for the first difference. We initially attempted to take this avenue, but as in Abrevaya the instruments were too imprecise.

shocks that are prevalent in an economic downturn are not necessarily the same as shocks occurring in a stable or growing economy.[10] For example, we should expect income shocks to be larger between 2000 and 2005, when Uruguay experienced a deep crisis (income fell in Uruguay from 1999 until 2002, and only recovered to pre-crisis levels in 2005). Furthermore, national trends in substance use show sizeable increases in the use of marijuana and cocaine after 2001, and discrete increases in the use of alcohol after 2006. We thus expect biases due to changes in substance use to be more critical in the 2000s than in the last years of the 1990s. To encompass these trends, we compare first-difference estimates for the periods 1995-1999, 2000-2005, and 2006-2011 and display them in Table 6. The first three columns compare prenatal care coefficients associated with the likelihood of preterm birth. The coefficients are fairly consistent for the 1995-1999 and 2006-2011 periods, and slightly higher for the 2000-2005 period (the recession and recovery timeframe). The last three columns report the effects of prenatal care on low birth weight without controlling for gestational age. As expected, we find larger coefficients for the period after 2000, suggesting potential biases due to income or substance use shocks. Still, the effects for 1995-1999 are large and statistically significant. Intermediate use of prenatal care decreases low birth weight by 3.2 percentage points and adequate use decreases it by 6 percentage points, relative to inadequate use.

Finally, we distinguish women who increased their use of prenatal care over two subsequent pregnancies from women who decreased their use of prenatal care. Our conjecture is that women who increase prenatal care use over time are likely to experience different types of shocks than women who reduce its use (that is, the shocks experienced by those reducing use are not necessarily the reversal of shocks that result in positive changes). Any difference could be attributed either to differential biases due to distinct types of shocks or to asymmetric effects. A finding of symmetric effects, on the other hand, would support the consistency of our estimates. The results of this comparison are shown in Table 7. While the coefficients are slightly higher for positive changers, the differences do not exceed 1 percentage point. This similarity in the estimates across women with opposing trends in the use of care is an additional indication of the robustness of our findings.

10. If these shocks occur and disappear at the same rate and if their effects are symmetric, they should not be a cause of concern. However, there may be periods in which the proportion of women experiencing these shocks may outnumber the fraction overcoming them (and vice versa).

Table 6. Effects of prenatal care (Kessner) on birth outcomes by delivery period

First-difference estimates

	Preterm birth			LBW No control for gestational age		
	1995-1999	2000-2005	2006-2011	1995-1999	2000-2005	2006-2011
	(1)	(2)	(3)	(4)	(5)	(6)
Intermediate prenatal care	-0.072***	-0.080***	-0.066***	-0.024***	-0.032***	-0.044***
	(0.012)	(0.007)	(0.011)	(0.007)	(0.006)	(0.009)
Adequate prenatal care	-0.097***	-0.120***	-0.103***	-0.057***	-0.060***	-0.065***
	(0.019)	(0.010)	(0.015)	(0.016)	(0.008)	(0.013)
N	4641	13463	5347	4641	13463	5347

Source: Authors' estimations. Data from the Perinatal Information System of Pereira Rossell Hospital (Ministry of Public Health).
Note: All estimations include the full set of controls depicted in Table 2 only for selected outcomes. Clustered standard errors in parentheses. * $p < 0.1$, ** $p < .05$, *** $p < .01$.

Table 7. Effects of prenatal care (Kessner) on birth outcomes, by sign of change in care use over time. First-difference estimates

Mothers with at least two deliveries and changes in prenatal care between 1995 and 2011

	Preterm birth		LBW No control for gestational age	
	Positive changers	Negative changers	Positive changers	Negative changers
	(1)	(2)	(3)	(4)
Intermediate prenatal care	-0.076***	-0.072***	-0.025**	-0.028**
	(0.013)	(0.013)	(0.011)	(0.011)
Adequate prenatal care	-0.108***	-0.104***	-0.043**	-0.040**
	(0.022)	(0.024)	(0.018)	(0.019)

Source: Authors' estimations. Data from the Perinatal Information System of Pereira Rossell Hospital (Ministry of Public Health).
Note: All estimations include the full set of controls depicted in Table 2. Clustered standard errors in parentheses. * $p < 0.1$, ** $p < .05$, *** $p < .01$.

5.3 Effectiveness of prenatal care over successive pregnancies

Our core estimates reflect average impacts of adequate vs. inadequate prenatal care on birth outcomes. An interesting question is whether this impact changes as the number of visits gets larger. Women could learn from past experiences and improve outcomes in future pregnancies, leaving lower margins for a direct effect of prenatal care in later pregnancies. We assess this question by rerunning the first-difference estimates only on first and second births, on second and third births, and on third and fourth births. Our findings (Table A2 of the appendix) confirm that the effects of prenatal care become smaller as the parity gets larger. Adequate prenatal care, according to the Kessner index, reduces the likelihood of preterm birth by 6.4 percentage points in the fourth pregnancy, versus a decrease of 11.8 percentage points in a second pregnancy. On the other hand, adequate prenatal care drops the likelihood of a low birth weight by 3.9 percentage points in a fourth pregnancy vs. 6.1 percentage points in a second pregnancy.

5.4 Marginal benefits of successive prenatal care visits

The analysis using the Kessner measures as indicators of adequacy of care has the advantage of adjusting prenatal care for the pregnancy's gestational length and for the timing of initiation of care. One drawback, however, is that it does not allow for a marginal evaluation of the benefits of successive visits. We explore this issue by running a semi-parametric regression of the number of prenatal care visits on birth outcomes, only for full-term pregnancies and conditioning on the number of gestational weeks to avoid endogenous mechanical relationships. The results are reported in Table A3 in the appendix. We find that prenatal care visits increase birth weight linearly, with changes of around 20 grams for every couple of visits. In contrast to the first four visits, the fifth and sixth visit contribute significantly to reducing the likelihood of low birth weight. Further visits continue to reduce the probability of low birth weight, but in smaller magnitudes.

6. Discussion and Concluding Remarks

In this paper we estimate the impact of prenatal care on infant health in a developing country by exploiting intra-mother variations in pregnancy inputs and outcomes. We analyze a longitudinal panel of births that took place between 1995 and 2011 in the largest public maternity ward in Uruguay, representative of the population of lower socioeconomic status in the country. The data set is quite unique in its ability to identify mothers throughout the period, its large size, and its reliance on clinical history (rather than self-reports).

Using within-mother first differences, we find that adequate use of prenatal care, as defined by early initiation and a minimum number of visits throughout the pregnancy, has a significant positive impact on neonatal outcomes. Our estimates indicate that the probability of low birth weight falls by between 3 and 6 percentage points (from a baseline prevalence of 10%), depending on the minimum number of control visits that are considered "adequate" (six or nine respectively), and the likelihood of preterm birth decreases by between 5 and 11 percentage points respectively (from a baseline of 14%).

By using first-differences estimation and difference-GMM estimation we avoid biases due to time-invariant, unobserved heterogeneity and feedback effects from prior births. We cannot claim causality, though, as these techniques do not account for time-variant shocks in preferences or health over different pregnancies. Still, we find our results to be strongly consistent when we run the analysis over different time periods (arguably subject to varying time shocks[11]) and when we restrict the estimation to mothers experiencing only positive variation or only negative variation in prenatal care use across pregnancies. Furthermore, one of the main sources of time-varying concerns–health shocks to the fetus or the mother–would bias our results towards zero.

Our estimates are larger than those obtained in other international studies using two-stage least squares (2SLS) and exploiting health policy changes as instruments in developed countries (Grossman and Joyce, 1990; Kaestner, 1999; Kaestner and Lee, 2003; Evans and Lien, 2005). We see two potential explanations for the discrepancy with prior studies. First, prenatal care may have a significantly larger effect among lower-SES women in a developing country than among women in a high-income country. In fact, our findings are in line with those of Wehby *et al.* (2009) for Argentina and Habibov and Fan (2011), Gajate-Garrido (2011), and Awiti (2014) for Azerbaijan, Cebu, and Kenya. Second, our statistical approach avoids the problems inherent in prior approaches that rely on instrumental variables to measure only local average treatment effects.

11. One of these shocks may have to do with the conditional cash transfer programs implemented in Uruguay after the 2002 crisis (PANES). Amarante *et al.* (2011) show that PANES had some positive impact on infant birth weight. However, the effects did not seem to be related to improved prenatal care.

One potential concern with our analysis has to do with biased attrition of mothers from the Pereira Rossell Hospital. Mothers who leave the sample (deciding to deliver their second, third, or fourth child at a hospital other than Pereira Rossell) may differ from mothers who remain in the sample. To address this concern, we explore changes in the fraction of nationwide births occurring at the Pereira Rossell Hospital between 1995 and 2011. We find two significant events that affected the number of deliveries at the hospital during the analysis period: the 2002 economic crisis and implementation of the health care reform in 2008. In 2002, when Uruguay experienced a severe economic crisis, the Pereira Rossell Hospital reached its peak coverage of 35% of all deliveries in Montevideo. Because only formal workers were entitled to receive care from private providers, the crisis shifted relatively better-off women from the private to the public sector. Once the economy began to recover in 2004, this fraction gradually returned to pre-crisis levels. The second important event was the implementation of the health care reform in 2008, which extended private health care coverage to spouses and dependents of formal workers. Two years after this event, births at Pereira Rossell Hospital fell to a minimum of 30% of births in Montevideo, the reform having shifted relatively better-off women out of the sample. While the impact of these sample variations is difficult to assess directly, the robustness of our estimates across different periods (Table 6) mitigates these concerns somewhat. In effect, prenatal care estimates are very similar across the 2000-2005 and the 2006-2011 periods, when attrition was working in opposite directions.

One of the reasons that our analysis focused on the Pereira Rossell Hospital is that this university hospital has good quality of information and a high level of coverage when compared to national birth registries, and it covers 70% of all births in the public sector in Montevideo. The external validity of our findings is closely related to the representativeness of this hospital's patients of the population of low-income mothers in Uruguay. We showed that women delivering at Pereira Rossell are slightly less educated and less likely to be married than other low-socioeconomic status women in the country, suggesting that our analysis applies to the poorest women in the country.

Several factors mediate the effectiveness of prenatal care on neonatal outcomes: the physician's influence on the woman's behavior during pregnancy (protein supplementation, recommendations to abstain from the use of alcohol, tobacco, and other drugs, among other behaviors), detection and treatment of conditions associated with low birth weight (syphilis, anemia, hypertension, urinary infections), the ability to detect early labor and prevent premature birth through bed rest and pharmacological agents, and preparation for delivery. In Uruguay, all prenatal care is provided by obstetricians. The usual paraclinical studies required in low-risk pregnancies are: 1) complete hemogram during the first visit and in the third trimester; 2) urine tests in all visits; 3) glycemia analysis in the first visit; 4) VDRL in the first visit and in the third trimester; 5) blood type and RH; 6) serology for toxoplasmosis; 7) testing for Hepatitis B antigen; 8) serology for Chagas disease; 9) HIV serology; 10) three obstetrical ultrasounds (one per trimester); 11) screening for gestational diabetes; 12) urine culture in the second and/or third trimester. Unfortunately, in this paper we are unable to disentangle the channels through which a prenatal care visit translates into better health outcomes at birth. Future research should explore these mechanisms.

Our findings have direct policy implications for developing countries, and particularly for the Uruguayan population. If the low-income population in Uruguay were to achieve the prenatal care goals set by the MPH, the impact on neonatal health would be substantial. However, only 22% of the population in our sample complied with the standards suggested by the MPH and only 13% displayed adequate use of prenatal care as defined by the Kessner criterion. These low figures, together with the magnitude of our findings, present a strong case for the design of policies aimed at encouraging the use of prenatal care in low-SES populations. The financial incentives that the Uruguayan Ministry of Health currently offers to health care providers for achieving a set of prenatal care goals may not be sufficient to promote significant changes among lower-income populations. Any policy aimed at improving prenatal care among low-SES women, either through financial incentives to providers or directly through conditional cash transfers or other demand-centered initiatives, must focus differentially on the segments of the population least likely to use prenatal care adequately, i.e. teenagers or older women, women who are single, uneducated, or those who have many children. Conditional cash transfer programs, which provide financial aid to low-income individuals upon compliance with health and education goals, should also align their incentives to elicit better compliance with prenatal care optimal standards.[12]

12. There is some evidence in Uruguay that the PANES, an emergency financial aid plan that took place after the 2002 crisis, had some positive impacts on infant birth weight (Amarante *et al.*, 2011). However, the effects did not seem to work through improved prenatal care.

REFERENCES

Abel, E.L., M. Kruger and L. Burd (2002), "Effects of maternal and paternal age on Caucasian and Native American preterm births and birth weights," *American Journal of Perinatology* **19**: 49–54.

Abrevaya, J. (2006), "Estimating the effect of smoking on birth outcomes using a matched panel data approach," *Journal of Applied Econometrics* **21**: 489–519.

Amarante, V., M. Manacorda, A. Vigorito and M. Zerpa (2011), "Social assistance and labor market outcomes: Evidence from the Uruguayan PANES," Inter-American Development Bank.

Arellano M. and S. Bond (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *Review of Economic Studies* **58**(2): 277–97.

Arellano M. and O. Bover (1995), "Another look at the instrumental variable estimation of error-components models," *Journal of Econometrics* **68**(1): 29–51.

Awiti, J. O. (2014) "A multilevel analysis of prenatal care and birth weight in Kenya," *Health Economics Review* **4**(1): 1–16.

Barros, F.C., Z.A. Bhutta, M. Batra, T.N. Hansen, C.G. Victora and C.E. Rubens (2010), "Global report on preterm birth and stillbirth (3 of 7): Evidence for effectiveness of interventions," *BMC Pregnancy Childbirth* 10, Suppl 1:S3.

Behrman, J.R. and M.R. Rosenzweig (2005), "Returns to birth weight," *Review of Economic and Statistics* **86**(2): 586–601.

Black, S., P. Devereux and K. Salvanes (2007), "From the cradle to the labor market? The effect of birth weight on adult outcomes," *Quarterly Journal of Economics* **122**(1): 409–39.

Boardman, J.D., D.A. Powers, Y.C. Padilla and R.A. Hummer (2002), "Low birth weight, social factors, and developmental outcomes among children in the United States," *Demography* **39**(2): 353–63.

Brien, M.J. and C.A. Swann (2001), "Prenatal WIC participation and infant health: Selection and maternal fixed effects." Manuscript, SUNY-Stony Brook Department of Economics.

Case, A., A. Fertig and C. Paxson (2004), "The lasting impact of childhood health and circumstance." Center for Health and Wellbeing, Woodrow Wilson School, Princeton University.

Centro Latinoamericano de Perinatología y Desarrollo Humano (CLAP-OPS/OMS), available at http://www.paho.org/clap/index.php?option = com_content&view = article&id = 84&Itemid = 242&lang = es.

Clement, S., *et al.* (1999), "Does reducing the frequency of routine antenatal visits have long term effects? Follow up of participants in a randomized controlled trial," *British Journal of Obstetrics and Gynecology* **106**(4), 367–70.

Conway, K.S. and P. Deb (2005), "Is prenatal care really ineffective? Or, is the 'devil' in the distribution?" *Journal of Health Economics* **24**(3): 489–513.

Currie, J. and B. Madrian (1999), "Health, health insurance, and the labor market," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Volume 3, 3309–415.

Currie, J. and E. Moretti (2005), "Biology as destiny? Short and long-run determinants of intergenerational transmission of birth weight" National Bureau of Economic Research Working Paper 11567.

Currie, J. and J. Grogger (2002), "Medicaid expansions and welfare contractions: Offsetting effects on prenatal care and infant health?" *Journal of Health Economics* **21**:313–36.

Ehrenberg, H.M., L. Dierker, C. Milluzzi and B.M. Mercer (2003. "Low maternal weight, failure to thrive in pregnancy, and adverse pregnancy outcomes," *American Journal of Obstetrics and Gynecology* **189**:1726–30.

Evans, W. N. and D.S. Lien (2005), "The benefits of prenatal care: Evidence from the PAT bus strike," *Journal of Econometrics* **125**(1): 207–39.

Figlio, D., S. Hamersma and J. Roth (2009), "Does prenatal WIC participation improve birth outcomes? New evidence from Florida" *Journal of Public Economics* **93**(1): 235–45.

Gajate-Garrido, G. (2013), "The impact of adequate prenatal care on urban birth outcomes: An analysis in a developing country context," *Economic Development and Cultural Change* **62**(1): 95–130.

Grossman, M. and T.J. Joyce (1990), "Unobservables, pregnancy resolutions, and birth weight production functions in New York City," *Journal of Political Economy* **98**: 983–1007.

Grossman, M. (2000) "The human capital model," in A. J. Culyer and J. P. Newhouse, eds., *Handbook of Health Economics* Vol. 1A. Amsterdam: North-Holland, Elsevier Science, 347–408.

Habibov, N.N. and L. Fan (2011), "Does prenatal healthcare improve child birthweight outcomes in Azerbaijan? Results of the National Demographic and Health Survey," *Economics & Human Biology* **9**(1): 56–65.

Haelterman, E., G. Breart, J. Paris-Llado, M. Dramaix and C. Tchobroutsky (1997), "Effect of uncomplicated ypertension on the risk of small-for-gestational age birth," *American Journal of Epidemiology* **145**: 689–95.

Holtz-Eakin, D., W. Newey and H.S. Rosen (1998), "Estimating vector autoregressions with panel data," *Econometrica* **56**: 1371–95.

Institute of Medicine (1985), *Preventing low birth weight.* Washington, D.C.: National Academy.

Jewell, R. T. (2007), "Prenatal care and birthweight production: Evidence from South America," *Applied Economics* **39**(4): 415–26.

Jewell, R.T. and P. Triunfo (2006), "The impact of prenatal care on birthweight: The case of Uruguay," *Health Economics* **15**: 1245–50.

Jewell, R.T., P. Triunfo and R. Aguirre (2007), "El peso al nacer de los niños en la principal maternidad del Uruguay: 1995 a 2004," *Revista Desarrollo y Sociedad* **59**:1–20, Universidad de los Andes, Colombia.

Joyce, T.J. and M. Grossman (1990), "Pregnancy wantedness and the early initiation of prenatal care," *Demography* **27**:1–17.

Kaestner, R. and W.C. Lee (2003), "The effect of welfare reform on prenatal care and birth weight," National Bureau of Economic Research Working Paper 9769.

Kaestner, R. (1999) "Health insurance, the quantity and quality of prenatal care, and infant health," *Inquiry* **36**(2): 162–75.

Khong, T.Y., E.D. Adema and J.J.H.M. Eerwich (2003), "On an anatomical basis for the increase in birth weight in second and subsequent born children," *Placenta* **24**: 348–53.

Kleinman, J.C. and S.S. Kessel (1987), "Racial differences in low birth weight: Trends and risk factors," *New England Journal of Medicine* **317**:749–53.

Kotelchuck, M. (1994), "An evaluation of the Kessner adequacy of prenatal care index and a proposed adequacy of prenatal care utilization index," *American Journal of Public Health* **84**: 1414–20.

McCormick, M.C. (1985), "The contribution of low birth weights to infant mortality and childhood morbidity," *New England Journal of Medicine* **312**: 82–90.

McDuffie, R., *et al.* (1996), "Effect of frequency of prenatal care visits on perinatal outcome among low-risk women," *The Journal of the American Medical Association* **275**: 847–51.

Ministry of Public Health (MPH) (2010), *Instructivo de metas asistenciales JUNASA_ MSP* (Health care goals manual), at *www.msp.gub.uy.*

Misra, D. and B. Guyer (1998), "Benefits and limitations of prenatal care: from counting visits to measuring content," *The Journal of the American Medical Association* **279**: 1661–2.

Permutt T. and J.R. Hebel (1989), "Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth weight," *Biometrics* **45**:619–22.

Petrou, S., T. Sach and L. Davidson (2001), "The long-term costs of preterm birth and low birth weight: Results of a systematic review," *Child: Care, Health and Development*, **27**: 97–115.

Reichman, N.E., H. Corman, K. Noonan and D. Dave (2009) "Infant health production functions: What a difference the data make?" *Health Economics* **18**(7): 761–82.

Roodman, D. (2009), "How to do xtabond2: An introduction to difference and system GMM in Stata," *Stata Journal* **9**: 86–136.

Rosenzweig, M.R. and T.P. Schultz (1982), "The behavior of mothers as inputs to child health: The determinants of birth weight, gestation, and rate of fetal growth," in V.R. Fuchs, ed., *Economic Aspects of Health.* Chicago: University of Chicago Press.

Rosenzweig, M.R. and T.P. Schultz (1983), "Estimating a household production function: Heterogeneity, the demand for health inputs, and their effects on birth weight," *Journal of Political Economy* **91**:723–46.

Rosenzweig, M.R. and T.P. Schultz (1988), "The stability of household production technology: A replication," *Journal of Human Resources* **23**: 535–49.

Thomas, P., J. Peabody, V. Turnier and R.H. Clark (2000), "A new look at intrauterine growth and the impact of race, altitude, and gender," *Pediatrics* **106**: E21.

Veloso da Veiga, P. and R. Wilder (2008), "Maternal smoking during pregnancy and birth weight: A propensity score matching approach," *Maternal and Child Health Journal* **12**: 194–203.

Villar, J., *et al.* (2001), "WHO antenatal care randomized trial for the evaluation of a new model of routine antenatal care," *Lancet* 357 (9268), 1551–1564.

Wehby G. L., J.C. Murray, E.E. Castilla, J.S. Lopez-Camelo and R.L. Ohsfeldt (2009), "Quantile effects of prenatal care utilization on birth weight in Argentina," *Health Economics* **18**(11): 1307–21.

APPENDIX

Table A1. Comparison of mothers in analysis sample (with at least two births in 1995-2011) with mothers in excluded sample (only one birth in the period)

Mother's characteristics at first observed pregnancy only

Dependent variables	Analysis sample N = 17,278	Excluded sample N = 65,136	Difference in means[a]
Preterm birth	0.145	0.147	
Birth weight	3110.29	3130.46	***
Very low birth weight	0.017	0.021	***
Low birth weight (< 2,500 grams)	0.107	0.116	***
Mother had adequate prenatal care (MPH)	0.232	0.335	***
Initiated prenatal care in first trimester	0.273	0.378	***
At least 6 prenatal care visits	0.512	0.634	***
Inadequate prenatal care (Kessner)	0.418	0.291	***
Intermediate prenatal care (Kessner)	0.439	0.489	***
Adequate prenatal care (Kessner)	0.142	0.220	***
Age < 17	0.098	0.066	***
Age 17-19	0.260	0.190	***
Age 20-34	0.596	0.625	***
Age 35-39	0.040	0.085	***
Age > 39	0.006	0.033	***
Marital status: cohabitation	0.179	0.199	***
Marital status: single	0.549	0.545	
Marital status: other	0.255	0.240	***
Marital status: married	0.017	0.017	
Education: did not finish primary school	0.157	0.103	***
Education: completed primary school	0.629	0.573	***
Education: completed middle school	0.186	0.266	***
Education: completed high school	0.028	0.057	***
Smoker	0.388	0.321	***
Mother is underweight	0.065	0.066	
Mother is overweight	0.071	0.095	***
Mother is obese	0.028	0.046	***
BMI missing	0.476	0.407	***
Previous stillbirth	0.025	0.025	
Previous abortions	0.200	0.239	***
Parity	1.423	1.458	**
First child	0.433	0.395	***

Table A1. (continued)

Dependent variables	Analysis sample N = 17,278	Excluded sample N = 65,136	Difference in means[a]
Hypertension	0.013	0.028	***
Hypertension missing	0.131	0.077	***
Preeclampsia	0.022	0.030	***
Preeclampsia missing	0.131	0.077	***
Eclampsia	0.002	0.002	
Eclampsia missing	0.133	0.078	***
Child's gender: Male	0.512	0.511	
Trimester of birth: 1	0.243	0.243	
Trimester of birth: 2	0.247	0.246	
Trimester of birth: 3	0.272	0.264	**

Source: Authors' estimations. Data from the Perinatal Information System of Pereira Rossell Hospital (Ministry of Public Health).
a. * $p < 0.1$, ** $p < .05$, *** $p < .01$.

Table A2. Effects of prenatal care (Kessner) on birth outcomes for different number of pregnancies

Selected outcomes

	Preterm Birth		
	1st-2nd pregnancies	2nd-3rd pregnancies	3rd-4th pregnancies
	(1)	(2)	(3)
Intermediate care	-0.081***	-0.076***	-0.034**
	(0.006)	(0.009)	(0.017)
Adequate care	-0.118***	-0.126***	-0.064**
	(0.009)	(0.013)	(0.030)
	Low Birth Weight No control for gestational age		
	1st-2nd pregnancies	2nd-3rd pregnancies	3rd-4th pregnancies
	(4)	(5)	(6)
Intermediate care	-0.034***	-0.026***	-0.026*
	(0.005)	(0.007)	(0.015)
Adequate care	-0.061***	-0.065***	-0.039*
	(0.007)	(0.011)	(0.023)

Source: Authors' estimations. Data from the Perinatal Information System of Pereira Rossell Hospital (Ministry of Public Health).
Note: All estimations include the full set of controls depicted in Table 2. Clustered standard errors in parentheses. * $p < 0.1$, ** $p < .05$, *** $p < .01$.

Table A3. Effects of prenatal care on birth outcomes for categorical ranges of prenatal care visits

Only full-term pregnancies and selected outcomes

	Preterm birth	LBW
	Adjusting for gestational age	
	(1)	(2)
1 and 2 visits	25.810**	-0.008
	(11.363)	(0.008)
3 and 4 visits	48.082***	-0.009
	(10.497)	(0.007)
5 and 6 visits	64.028***	-0.019***
	(10.483)	(0.007)
7 and 8 visits	84.155***	-0.034***
	(10.780)	(0.007)
9 and 10 visits	117.884***	-0.036***
	(11.365)	(0.007)
11 + visits	208.778***	-0.052***
	(26.053)	(0.016)

Source: Authors' estimations. Data from the Perinatal Information System of Pereira Rossell Hospital (Ministry of Public Health).
Note: All estimations include the full set of controls depicted in Table 2. Clustered standard errors in parentheses. * $p < 0.1$, ** $p < .05$, *** $p < .01$.

VOL. 52 NO. 2 (NOV, 2015), 185–211

LATIN AMERICAN JOURNAL OF ECONOMICS

APPLICATION OF A SHORT MEMORY MODEL WITH RANDOM LEVEL SHIFTS TO THE VOLATILITY OF LATIN AMERICAN STOCK MARKET RETURNS*

GABRIEL RODRÍGUEZ**
ROXANA TRAMONTANA TOCTO***

Empirical research indicates that the volatility of stock return time series has long memory. However, it has been demonstrated that short memory processes contaminated by random level shifts can often be confused with long memory, a feature often referred to as spurious long memory. This paper represents an empirical study of the random level shift (RLS) model for the volatility of daily stock return data for five Latin American countries. This model consists of the sum of a short term memory component and a level shift component that is governed by a Bernoulli process with a shift probability α. The results suggest that level shifts in the volatility of daily stock return data are infrequent but when taken into account, the long memory characteristic and GARCH effects disappear. An out-of-sample forecasting exercise is also provided.

JEL classification: C22

Keywords: Volatility, long memory, random level shifts, forecasting, Latin America

1. INTRODUCTION

There are two important stylized facts that are found in returns from financial market variables such as stocks and exchange rates: the long-memory behavior of the volatility of returns and the presence of GARCH effects. Fractionally integrated processes have become a standard class of models used to describe the long-memory features of economic and financial time series data. Let $\{y_t\}_{t=1}^{T}$ be a stationary time series. Let $\gamma_y(\tau)$ be the autocovariance function of y_t, so y_t has

* The authors are grateful for useful comments from Paul Castillo (Central Bank of Peru), Rodolfo Cermeño (CIDE), Jiawen Xu (Shangai University of Finance and Economics), Zhongjun Qu and Pierre Perron (Boston University), and Patricia Lengua Lafosse (PUCP). We also appreciate the helpful comments from participants at the 32nd Meeting of Economists hosted by the Central Reserve Bank of Peru in November 2014. Finally, we would like to acknowledge the observations made by editor Felipe Zurita and an anonymous referee. All remaining errors are our responsibility.

** Corresponding author. Department of Economics, Pontificia Universidad Católica del Perú, Av. Universitaria 1801, Lima 32, Lima, Peru. Telephone: +511-626-2000 (4998). Email: gabriel.rodriguez@pucp.edu.pe.

** This paper is drawn from the thesis of Roxana Tramontana Tocto, Department of Economics, Pontificia Universidad Católica del Perú.

doi 10.7764/LAJE.52.2.185

long memory if $\gamma_y(\tau) = c(\tau)\tau^{2d-1}$, for $\tau \Rightarrow \infty$, where $c(\tau)$ is a smooth variation function. It implies that the autocorrelation function (ACF) decays to a hyperbolic rate[1]. On the other hand, $\{y_t\}_{t=1}^{T}$ has spectral density function $f_y(\omega)$ in the frequency ω, so y_t has long memory if $f_y(w) = g(\omega)\omega^{-2d}$, for $\omega \Rightarrow 0$; where $g(\omega)$ is a smooth variation function in a neighborhood of the origin, which means that for all real numbers t, it is verified that $g(t\omega)\ /\ g(\omega) \Rightarrow 1$ for $w \Rightarrow 0$. When $d > 0$, the spectral density function is growing for frequencies that are increasingly close to the origin. The rate of divergence to infinity depends on the given value of the parameter d.

A vast literature exists in estimating the long memory parameter d. Granger and Joyeux (1980) developed for the first time the notion of fractional integration in terms of an infinite filter corresponding to the expansion of $(1 - L)^d$, where L is the lag operator. When this expansion is applied to a white noise, we get a series with long memory. Then Hosking (1981) developed the ARFIMA(p,d,q) model that generalizes the autoregressive integrated moving average processes incorporating fractional values for the integration parameter d. When $0 < d < 0.5$, the fractional integration process has long memory, and when $-0.5 < d < 0.5$ the series is stationary. Geweke and Porter-Hudak (1983) show that the asymptotic distribution of the integration parameter d has a normal distribution based on a linear regression of the log-periodogram with a deterministic regressor; see also Robinson (1995).

In the context of GARCH models, Baillie *et al.* (1996) proposed the FIGARCH model, where the fractional integration parameter determines that shocks to the conditional variance disappear at a hyperbolic rate of decay. This characteristic explains the temporal dependencies in financial market volatility. Bollerslev and Mikkelsen (1996) extended the model to the fractional integration exponential GARCH (FIEGARCH). In both cases, the fractional parameter is significant and asymmetries are identified. Ding *et al.* (1993) estimate the ARCH model, taking into account the squared returns and absolute returns, and show the existence of long memory. Then the authors propose the asymmetric power GARCH model (APARCH), allowing the long memory parameter in the volatility and the asymmetry parameter. Finally, Lobato and Savin (1998) apply a semiparametric test to detect the presence of long memory in the daily S&P 500 stock returns and their squares. The

1. A practical definition of long memory is to state that the sum of the autocorrelations is infinite; that is, $\lim_{T\Rightarrow\infty}\Sigma_{j=-T}^{T}|\rho_j|=\infty$.

short memory null hypothesis is not rejected for the stock returns while the null hypothesis is rejected for the squared and absolute returns. However, the authors argue that their results could be spurious for the squared stock returns due to the nonstationarity of the series and for the absolute values due to the aggregation.

Studies have demonstrated that structural break processes and processes with non-linear features are often confused with long memory, a feature that is frequently referred to as spurious long memory. A steadily growing literature has developed that emphasizes whether it is possible to empirically discriminate between true long memory processes (or fractionally integrated processes) and spurious long memory processes. Perron (1989, 1990) has shown that when there is a contaminated stationary process with structural breaks, the sum of the autoregressive coefficients is biased to the unit. Diebold and Inoue (2001) show that the change of stochastic regime is easily confused with long memory, even asymptotically, provided that the probabilities of structural breaks are small; see also Engle and Smith (1999). Using Monte Carlo simulation, they emphasize the relevance of the theory in finite samples and make it clear that the confusion is not merely a theoretical question, but a real possibility in empirical applications in economic and finance time series. Gourieroux and Jasiak (2001) find that nonlinear time series with infrequent linear breaks could have long memory on the basis of the estimation of the correlogram instead of estimation of the fractional parameter. These findings show that these series, rather than the fractionally integrated processes with *i.i.d.* innovations, would generate the hyperbolic decay of the autocorrelogram. Granger and Hyung (2004) show that the slow decay in autocorrelation and other properties of fractionally integrated models are generated by occasional breaks. The authors find that not taking the breaks into account generates the presence of long memory in the ACF and the fractional parameter estimated using the method of Geweke and Porter-Hudak (1983) is biased. Mikosch and Stărică (2004a) provide a theoretical basis to explain two stylized facts that are observed in the logarithm of returns: long-range dependence in volatility and the integrated GARCH (IGARCH) if it is assumed that the data are not stationary. The simulations show that the time series with changing unconditional variance produce estimates of the long memory parameter d that may be erroneously interpreted as evidence of long memory under the assumption of stationarity. There is evidence that the characteristic of long-range dependence is caused by feasible structural changes in the logarithm of stock market returns.

Also, Mikosch and Stărică (2004b) propose a goodness of fit test that shows the similarity between the spectral density of a GARCH process and the logarithm of stock market returns that detect changes in the structure of the data that are related to changes in the unconditional variance. These changes would induce long-range dependence in the ACF of absolute returns; see also Stărică and Granger (2005).

Perron and Qu (2010) perform an analysis of various statistics when the underlying model is a short memory process with random level shifts rather than a fractionally integrated process. They analyze the estimates of the ACF, the periodogram, and the log-periodogram. The results show that a short memory process with level shifts is a good candidate for modeling volatility. The estimates clearly follow a pattern that would be obtained if the underlying process was short memory with level shifts. Lu and Perron (2010) estimate a random level shift (RLS) model that consists of the sum of a short-term process and a level shift component, where the shift component is governed by a Bernoulli process with a shift probability α. The estimation method transforms the model into linear state-space equations with a mixture of normal innovations to apply the Kalman filter. The results show that there is limited evidence of correlation in the remaining noise; therefore, there is no evidence of long memory. On the other hand, once the level shifts found are introduced into a GARCH model and applied to the series, any evidence of GARCH effects disappears. For predictions outside the sample of squared returns, in most cases the RLS model performs better than the GARCH (1,1) model and the fractionally integrated GARCH. Similar results are found by Li and Perron (2013) but applied to two exchange rates.

Empirical studies for financial variables in Latin America are scarce. A research agenda for the region was suggested in Humala and Rodríguez (2013), of which this work forms part. The main aim of this paper is to estimate an RLS model to the volatilities of returns of five Latin American financial markets following the approach of Lu and Perron (2010) and Li and Perron (2013). The results show that the probability of level shifts is small but they are responsible for the presence of long memory in the volatilities of the series analyzed. Having estimated the probability of level shifts, the exact number of such level shifts can be calculated. Thus, the component obtained as a difference between the volatility series and the level shifts possesses an ACF that indicates an absence of long memory. We show that short memory processes contaminated by random level shifts can be confused with long memory in the data considered. Finally, an out-of-sample forecasting exercise shows that the RLS model performs better than traditional models for modeling long memory, such as the ARFIMA (p,d,q) models.

This paper is structured as follows: Section 2 presents the RLS model and some details related to the estimation algorithm. Section 3 presents the empirical results, which are divided into two aspects: the effects of the level shift component on long memory and on GARCH components. Section 4 discusses the performance of the RLS in terms of prediction, while Section 5 presents concluding remarks.

2. The Model

We follow the approach and notation of Lu and Perron (2010) by using a simple mixture model that is a combination of a level-shift component depending on a Bernoulli distribution and a short-memory process. The RLS model is specified as follows:

$$\begin{aligned} y_t &= a + \tau_t + c_t, \\ \tau_t &= \tau_{t-1} + \delta_t, \\ \delta_t &= \pi_t \eta_t, \end{aligned} \tag{1}$$

where a is a constant, τ_t is the level shift component and c_t is the short memory process, while π_t is a Bernoulli variable, which takes the value of 1 with probability α and the value of 0 with probability $(1-\alpha)$. Thus, in the third expression in (1), when π_t assumes the value of 1, a random level shift η_t occurs with a distribution $\eta_t \sim i.i.d.N(0,\sigma_\eta^2)$. The short memory process (in its general form) is defined by the process $c_t = C(L)\, e_t$, with $e_t \sim i.i.d.N(0,\sigma_e^2)$ and $E|e_t|^r < \infty$ for values $r > 2$ and where $C(L) = \sum_{i=0}^{\infty} c_i L^i$, $\sum_{i=0}^{\infty} i|c_i| < \infty$ and $C(1) \neq 0$. Likewise, it is assumed that π_t, η_t and c_t are mutually independent. Based on the results of Lu and Perron (2010) and Li and Perron (2013), even though it would be worthwhile to consider the component e_t as a random variable (noise), in this paper we model this component as an AR(1) process; that is, $c_t = \phi c_{t-1} + e_t$[2].

2. We opted for an AR(1) specification but if the coefficient ϕ is statistically insignificant, $c_t = e_t$. Estimates with longer lags for the AR process showed no significance of the respective parameters. This is consistent with the results of the RLS model because if persistence or long memory in the volatility of the series analyzed is mainly explained by rare or sporadic level shifts, then the short-memory component contains little persistence or is a noise. This justifies modeling c_t as a noise or not more than an AR(1) process.

In comparison with Hamilton's Markov-switching model (1989), this model does not limit the magnitude of level shifts, meaning that any number of patterns is possible. Moreover, the probability 0 or 1 does not depend on past facts, unlike the Markov model. Note that the process δ_t can be described as $\delta_t = \pi_t \eta_{1t} + (1 - \pi_t)\eta_{2t}$ with $\eta_{it} \sim i.i.d.N(0,\sigma^2_{\eta_i})$ for $i = 1, 2$ and $\sigma^2_{\eta_1} = \sigma^2_\eta$, $\sigma^2_{\eta_2} = 0$. To eliminate the autoregressive process of the level shift component, the first difference model only depends on the Bernoulli process: $\Delta y_t = \tau_t - \tau_{t-1} + c_t - c_{t-1} = c_t - c_{t-1} + \delta_t$, and passing to the state space form, the measurement and transition equations are obtained, respectively: $\Delta y_t = c_t - c_{t-1} + \delta_t$, $c_t = \phi c_{t-1} + e_t$. In matrix form we have $\Delta y_t = HX_t + \delta_t$ and $X_t = FX_{t-1} + U_t$, where $X_t = [c_t, c_{t-1}]$, $F = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix}$, $H = [1, -1]'$. In this case, the first row of the matrix F shows the coefficient ϕ of the autoregressive part of the short memory component. Moreover, U is a normally distributed vector of dimension 2 with mean 0 and variance: $Q = \begin{bmatrix} \sigma^2_e & 0 \\ 0 & 0 \end{bmatrix}$. In comparison with the standard state space model, the major difference in the current model is that the distribution of δ_t is a mixture of normal differences with variance σ^2_η and 0, occurring with probabilities α and $1 - \alpha$, respectively[3].

The model set out above is a special case of the models employed in Wada and Perron (2006) and Perron and Wada (2009). In this case, there are only shocks that affect the level of the series, with the restriction imposed that the variance of one of the components of the mixture of distributions is zero. The basic input for the estimation is the increase in the states through realizations of the mixture at time t so that the Kalman filter can be used to form the conditional likelihood function for the realizations of the states. The latent states are eliminated from the final likelihood expression by adding in all possible realizations of the states. As a consequence, despite its fundamental differences, the model takes a structure that is similar to Hamilton's Markov-switching model (1994). Let $Y_t = (\Delta y_1, ..., \Delta y_t)$ be the vector of available observations at time t and the vector of parameters is denoted by $\theta = [\sigma^2_\eta, \alpha, \sigma^2_e, \phi]$. Adopting the notation used in Hamilton (1994), $\mathbf{1}$ represents a (4×1) vector of ones, the symbol $\odot$ denotes element-by-element multiplication, $\hat{\xi}^{ij}_{t|t-1} = vec(\tilde{\xi}_{t|t-1})$ with the (i,j)th element of $\tilde{\xi}_{t|t-1}$ being $\Pr(s_{t-1} = i,\ s_t = j \mid Y_{t-1}; \theta)$ and $\omega_t = vec(\tilde{\omega}_t)$ with the (i,j)th element of $\tilde{\omega}_t$ being $f(\Delta y_t \mid s_{t-1} = i,\ s_t = j,\ Y_{t-1}; \theta)$ for $i, j \in \{1, 2\}$. Thus, we have $s_t = 1$ when $\pi_t = 1$; that is, a level

3. Note that this model could be extended to model the short memory component as an ARMA process.

shift occurs. Using the same notation as Lu and Perron (2010), the logarithm of the likelihood function is $\ln(L) = \Sigma_{t=1}^{T} \ln f(\Delta y_t \,|Y_{t-1}; \theta)$, where $f(\Delta y_t \,|Y_{t-1}, \theta) = \Sigma_{i=1}^{2} \Sigma_{j=1}^{2} f(\Delta y_t \,|s_{t-1} = i, s_t = j, Y_{t-1}, \theta) \Pr(s_{t-1} = i, s_t = j \,|\, Y_{t-1}, \theta) \equiv \mathbf{1}' \,\hat{\xi}_{t|t-1} \odot \omega_t$. By applying conditional probability rules, Bayes' rule and the independence of s_t with respect to past realizations, we have $\tilde{\xi}_{t|t-1}^{ki} = \Pr(s_{t-2} = k,\, s_{t-1} = i \,|\, Y_{t-1}; \theta)$. The evolution of $\hat{\xi}_{t|t-1}$ can be expressed as:

$$\begin{bmatrix} \tilde{\xi}_{t+1|t}^{11} \\ \tilde{\xi}_{t+1|t}^{21} \\ \tilde{\xi}_{t+1|t}^{12} \\ \tilde{\xi}_{t+1|t}^{22} \end{bmatrix} = \begin{bmatrix} \alpha & \alpha & 0 & 0 \\ 0 & 0 & \alpha & \alpha \\ 1-\alpha & 1-\alpha & 0 & 0 \\ 0 & 0 & 1-\alpha & 1-\alpha \end{bmatrix} \begin{bmatrix} \tilde{\xi}_{t|t}^{11} \\ \tilde{\xi}_{t|t}^{21} \\ \tilde{\xi}_{t|t}^{12} \\ \tilde{\xi}_{t|t}^{22} \end{bmatrix},$$

which is equal to $\hat{\xi}_{t+1|t} = \Pi \hat{\xi}_{t|t}$ with $\hat{\xi}_{t|t} = ((\hat{\xi}_{t|t-1} \odot \omega_t)/(\mathbf{1}' \,\hat{\xi}_{t|t-1} \odot \omega_t))$. As a consequence, the conditional likelihood function for Δy_t follows the following normal density:

$$\begin{aligned} \tilde{\omega}_t^{ij} &= f(\Delta y_t \,|\, s_{t-1} = i, s_t = j, Y_{t-1}, \theta) \\ &= \frac{1}{\sqrt{2\pi}} \left| f_t^{ij} \right|^{-1/2} \exp\left(-\frac{v_t^{ij'} (f_t^{ij})^{-1/2} \, v_t^{ij}}{2} \right), \end{aligned} \tag{2}$$

where v_t^{ij} is the prediction error and f_t^{ij} is its variance, and these terms are defined as: $v_t^{ij} = \Delta y_t - \Delta y_{t|t-1}^{i} = \Delta y_t - E[\Delta y_t | s_t = i, Y_{t-1}; \theta]$ and $f_t^{ij} = E(v_t^{ij} v_t^{ij'})$. The best predictions for the state variable and its respective conditional variance in $s_{t-1} = i$ are $X_{t|t-1}^{i} = FX_{t-1|t-1}^{i}$ and $P_{t|t-1}^{i} = FP_{t-1|t-1}^{i} F' + Q$, respectively.

The measurement equation is $\Delta y_t = HX_t + \delta_t$, where the error δ_t has mean 0 and a variance that can take values $R_1 = \sigma_\eta^2$ with probability α or $R_2 = 0$ with probability (1–α). Thus, the prediction error is $v_t^{ij} = \Delta y_t - HX_{t|t-1}^{i}$ and its variance is $f_t^{ij} = HP_{t|t-1}^{i} H' + R_j$. In this way, given that $s_t = j$ and $s_{t-1} = i$ and using update formulas, we have $X_{t|t}^{ij} = X_{t|t-1}^{i} + P_{t|t-1}^{i} H'(HP_{t|t-1}^{i} H' + R_j)^{-1}(\Delta y_t - HX_{t|t-1}^{i})$ and $P_{t|t-1}^{ij} = P_{t|t-1}^{i} - P_{t|t-1}^{i} H'(HP_{t|t-1}^{i} H' + R_j)^{-1} HP_{t|t-1}^{i}$. In order to reduce the dimensionality problem in the estimation, Lu and Perron (2010) use the re-collapsing procedure suggested by Harrison and Stevens (1976). In doing so, ω_t^{ij} is unaffected by the history of the states before time $t-1$. We have four possible states corresponding to $S_t = 1$ when $(s_t = 1,\ s_{t-1} = 1)$, $S_t = 2$ when $(s_t = 1,\ s_{t-1} = 2)$, $S_t = 3$ when $(s_t = 2,\ s_{t-1} = 2)$ and $S_t = 4$ when $(s_t{=}2,\ s_{t-1} = 2)$ and the matrix Π is defined as before. Taking the definitions of ω_t, $\hat{\xi}_{t|t}$, $\hat{\xi}_{t+1|t}$, the group of conditional probabilities, and one-period forward predictions, the same structure as a version of Hamilton's Markov model (1994) is obtained. Nonetheless, the EM algorithm cannot be utilized. This is because the mean and the variance in the function of conditional density are non-linear functions of the parameters θ and of the past realizations $\{\Delta y_{t-j}; j \geq 1\}$. Likewise, the conditional probability of being in a given regime $\hat{\xi}_{t|t}$, is not separable from the conditional densities ω_t. For further details, see Lu and Perron (2010), Li and Perron (2013), and Wada and Perron (2006).

Once the point estimate of α is obtained, a possible path is the use of a smoothed estimate of the level shift component τ_t. Nonetheless, in this context of abrupt structural changes, the conventional smoothers perform poorly. Instead, we use the method proposed by Bai and Perron (1998, 2003) to obtain the dates on which the level shifts occur as well as the means (averages) inside each segment. Thus, we use the estimate α to obtain an estimate of the number of level shifts and the Bai and Perron method (1998, 2003) to obtain the estimates of break dates that globally minimize the following squared residuals: $\sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} [y_t - \mu_i]^2$, where m is the number of breaks, T_i $(i = 1,2,...; m)$ are the break dates with $T_0 = 0$, and $T_{m+1} = T$ and μ_i $(i = 1,2,...,\ m + 1)$ are the means (averages) inside each regime, which can be estimated once the break dates are estimated. This method is efficient and can manage a large number of observations; see Bai and Perron (2003) for further details[4].

3. Empirical Results

We use five daily series that are the volatilities of the returns of the major Latin American stock markets: MERVAL (Argentina) from January 5, 1988 to June 13, 2013 (6,284 observations); IBOV (Brazil)

4. Note that since the model allows for consecutive level shifts, we set the minimum segment length to only one observation.

from January 2, 1992 to June 13, 2013 (5,303 observations); IPSA (Chile) from January 3, 1989 to June 13, 2013 (6,097 observations); MEXBOL (Mexico) from January 20, 1994 to June 13, 2013 (4,840 observations), and IGBVL (Peru) from January 3, 1990 to June 13, 2013 (5,832 observations). The returns series are generated as $r_t = \ln(P_t) - \ln(P_{t-1})$, where P_t is the closing price index of the respective stock market. Following recent literature (see Lu and Perron (2010) and Li and Perron (2010), among others), we model log-absolute returns.[5]. When returns are zero or close to it, the log-absolute transformation implies extreme negative values. Using the estimation method described in Section 2, these outliers would be attributed to the level shifts component and thus bias the probability of shifts upward. To avoid this inconvenience, we bound absolute returns away from zero by adding a small constant, i.e., we use $y_t = \log(|r_t| + 0.001)$, a technique introduced to the stochastic volatility literature by Fuller (1996).[6] The results are robust to alternative specifications, such as using another value for this so-called offset parameter, deleting zero observations, or replacing them with a small value. It is also important to note that we use daily returns as opposed to realized volatility series constructed from intra-daily high-frequency data, a method that has recently become popular. While realized volatility series are a less noisy measure of volatility, they are problematic in the current context for the following reasons: (i) such series are typically available for short span. Given that the level shifts will be relatively rare, it is essential to have data covering a long span of time in order to build reliable estimates of the probability of occurrence of the level shifts; (ii) such series are available only for specfic assets as opposed to market indices. Because the goal of the RLS model is to allow for particular events that affect overall markets, using a specific asset would confound such market-wide events with idiosyncratic ones associated with the particular asset used; and (iii) we are interested in re-evaluating the adequacy of ARFIMA and GARCH models applied to daily returns when taking into account the possibility of level shifts. Therefore, it is important to have estimates of these

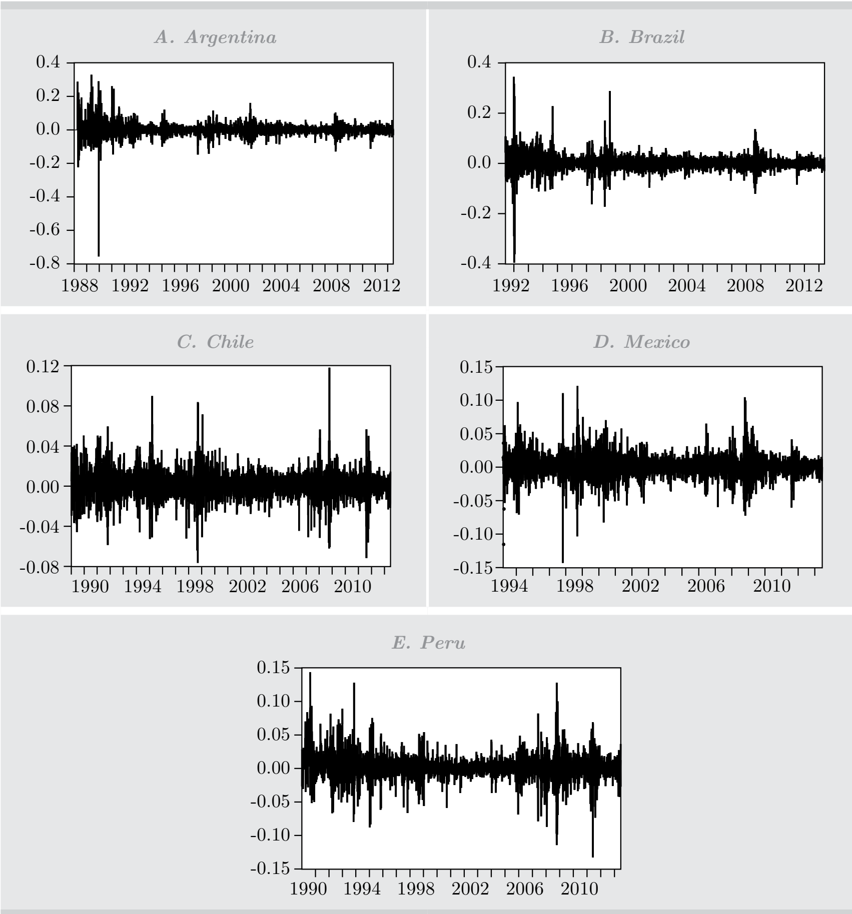
5. Using this measure has two advantages: (i) it does not suffer from a non-negativity constraint as do absolute or squared returns, for example. Actually, it is a similar argument as that used in the EGARCH(1,1) model proposed by Nelson (1991): the dependent variable is $\log(\sigma_t^2)$ in order to avoid the problems of negativity when the dependent variable is σ_t^2 as in the standard GARCH models and other relatives models; (ii) there is no loss relative to using square returns in identifying level shifts since log-absolute returns is a monotonic transformation. It is true that log-absolute returns are quite noisy but this is not problematic since the algorithm used is robust to the presence of noise.

6. All data have been obtained from Bloomberg.

level shifts for squared daily returns, which are equivalent to those estimated using log-absolute returns.

Figure 1 shows the returns series for five economies, while Table 1 shows the descriptive statistics of the returns and volatilities. In Figure 1, the frequent high-variation grouping of returns during periods of international or local crisis can be observed. The values of the descriptive statistics show a mean close to zero. On the other hand, the standard deviation is different, and the markets of Argentina, Brazil, and Peru are the most volatile. Figure 2 shows the ACF for the five series for 2,000 lags. In all cases, the long memory evidence is clear.

Figure 1. Stock returns series



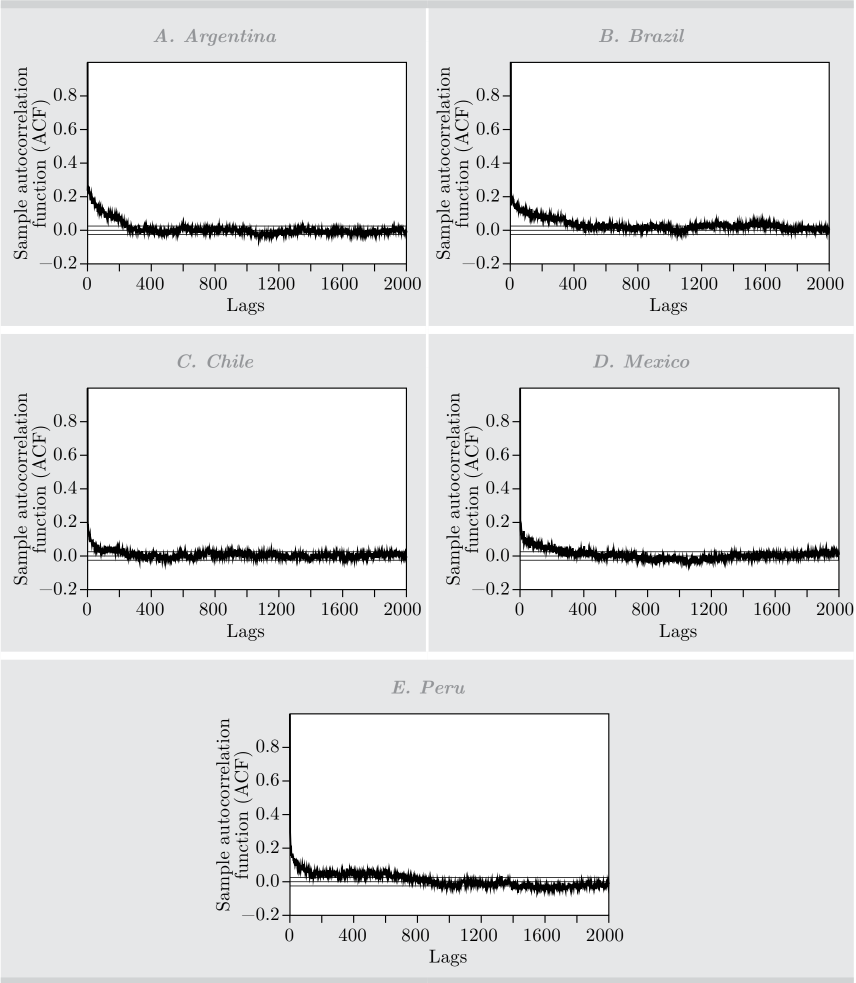
Source: Authors' calculations.

Table 1. Descriptive statistics

Country	Mean	SD	Max.	Min.	Skewness	Kurtosis	Jarque-Bera	Prob.
				Returns				
Argentina	0.002	0.033	0.329	-0.757	-0.574	56.744	756638.8	0.000
Brazil	0.003	0.046	0.693	-0.693	-0.703	105.558	2544970	0.000
Chile	0.000	0.012	0.118	-0.077	-0.182	8.694	8270.115	0.000
Mexico	0.001	0.016	0.122	-0.143	-0.020	9.595	8770.366	0.000
Peru	0.001	0.017	0.143	-0.132	-0.519	11.088	16159.29	0.000
				Volatility				
Argentina	-4.434	1.112	-0.277	-6.908	-0.271	2.799	87.557	0.000
Brazil	-4.502	1.173	-0.365	-6.901	-0.275	2.995	73.195	0.000
Chile	-4.992	0.845	-2.128	-6.908	-0.151	2.538	77.484	0.000
Mexico	-4.797	0.894	-1.937	-6.908	-0.161	2.604	52.561	0.000
Peru	-4.858	0.951	-1.931	-6.908	-0.027	2.622	35.376	0.000

Source: Authors' calculations.

Figure 2. Sample ACF of returns volatility series



Source: Authors' calculations.

3.1 Effects of level shifts on long memory and ARFIMA models

The estimated parameters are set out in Table 2[7] and correspond to the standard deviation of the level shift component σ_η, the probability of a level shift α, the standard deviation of the stationary component σ_e and the autoregressive ϕ of the specification $AR(1)$ for c_t. All estimated coefficients can be seen to be significant.

Table 2. Estimates of the RLS model

Country	$\sigma\eta$	α	σ_e	ϕ	**Likelihood**
Argentina	1.122^a (0.131)	0.004^a (0.001)	0.964^a (0.009)		8868.144
Brazil	0.425^a (0.118)	0.010^a (0.006)	0.881^a (0.009)		6965.685
Chile	0.612^a (0.150)	0.008^a (0.004)	0.778^a (0.007)	0.080^a (0.014)	7299.642
Mexico	0.520^a (0.157)	0.006^a (0.004)	0.830^a (0.009)	0.025^a (0.015)	6067.754
Peru	0.875^a (0.128)	0.004^a (0.002)	0.842^a (0.008)	0.115^a (0.015)	7421.854

Source: Authors' calculations.
Standard errors are in parentheses; *a*, *b*, *c* denote significance at 1%, 5%, and 10%, respectively.

The estimate of ϕ is not significant for the cases of Argentina and Brazil. In the other cases, although it is significant, it is nevertheless small. On the other hand, the probability of level shift is small in all of the cases considered. Therefore, given T and the estimates of α, we can find the number of level shifts for each country: Argentina has 25, Brazil has 53, Chile has 49, Mexico has 29, and Peru has 26. These values indicate that level shifts are rare and have a duration of 222, 98, 124, 161 and 216 days on average for Argentina, Brazil, Chile, Mexico and Peru, respectively[8].

7. Given that all components of the vector of states are stationary, we initialize the vector of states and its variance matrix through its unconditional expected values: $X_{0|0} = (0{,}0)'$ and $P_{0|0} = \begin{bmatrix} \sigma_e^2 & 0 \\ 0 & 0 \end{bmatrix}$. In order to avoid the problem of local maximums, we re-estimate the model using a long group of random initial values and select the related estimates with the highest likelihood after finding convergence.

8. Observing the distribution of the level shifts, we find that the smallest occurrence of level shifts are 2, 1, 3, 4, and 3 days for Argentina, Brazil, Chile, Mexico and Peru, respectively. Their maximum values are, respectively: 1,010, 750, 614, 734, and 1,715 days.

Figure 3 presents the smoothed (Gaussian kernel) series of the level shift component and the level shift component with dates estimated using the method of Bai and Perron (1998, 2003). The smoothed estimates are erratic, even though they closely follow changes in the mean of the series as indicated by the method of Bai and Perron (1998, 2003). Figure 3 shows the grouping of the dates of level shifts where volatility has experienced strong variations in short periods of time. Figure 3 also shows that the main dates of level shifts are similar across the five countries. These level shifts proceed from two sources. The first is external: crises occurring in other countries, including the Asian, Russian and Mexican crises that affected the volatility of the five markets analyzed. Another important external factor was the international financial crisis (2008) that affected all of the economies analyzed. The second is domestic: level shifts caused by primary election periods, monetary crises that trigger periods of high inflation, as well as societal events. All these factors contribute to the presence of level shifts in the volatility series. Argentina and Brazil experience successive level shifts at the start of their samples due to high inflationary processes that have occurred in these countries. The continuous level shifts in Brazil continue until 1998. The greatest level shifts in Chile are discerned in 1990-1996. In the case of Peru, level shifts in 1990, 1996, and 1998 are appreciated[9]. The year 2008 clearly affects all economies.

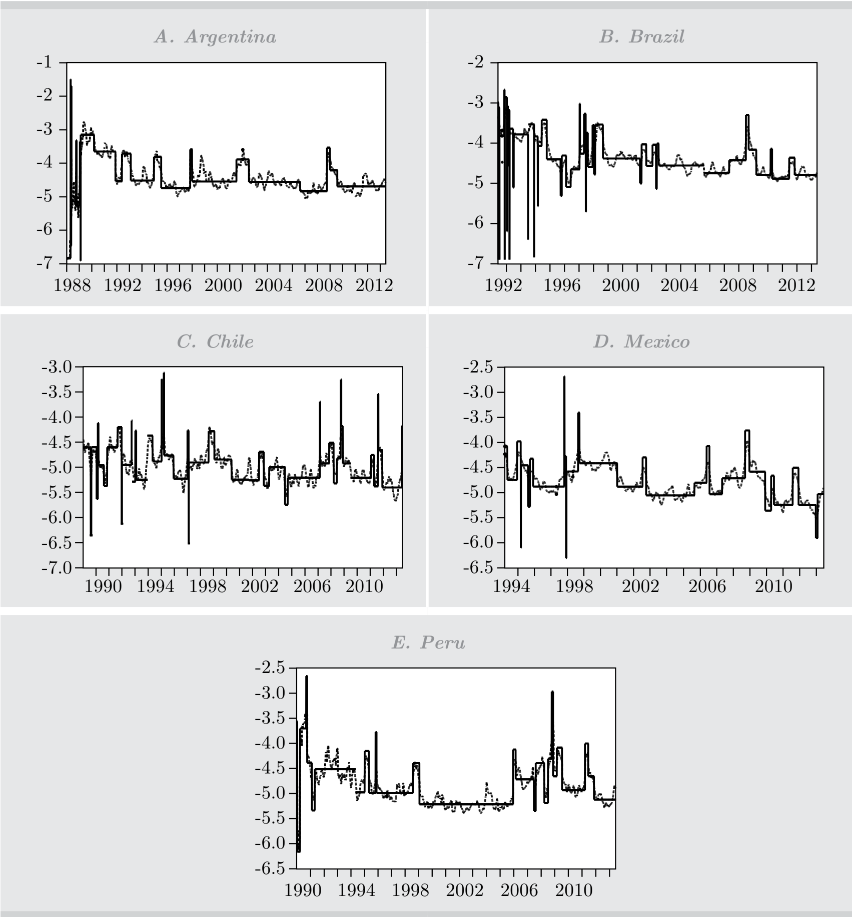
The ACF was estimated by leaving aside the estimated level component using the algorithm of Bai and Perron (2003). What is observed (Figure 4) is that all traces of long memory have disappeared[10]. A complementary way of analyzing the long-memory characteristic is by estimating the ARFIMA(0,d,0) and ARFIMA (1,d,1) models for the volatility series and the volatility series excluding the level component[11]. In the case of the ARFIMA(0,d,0) estimates, the results convey the same message observed in the ACF in Figure 4. The estimates of the fractional parameter (d) indicate long-term behavior, which is a stylized fact frequently mentioned in the literature. However, the short-term component shows fractional parameter estimates $\widehat{d}$ with values that are positive but too small to imply long memory or in other cases the values are negative, implying anti-persistence. The results are similar

9. A more detailed account of the Peruvian case can be found in Ojeda Cunya and Rodríguez (2015).

10. Similar results are obtained when we use the smoothed estimates of the level shift component.

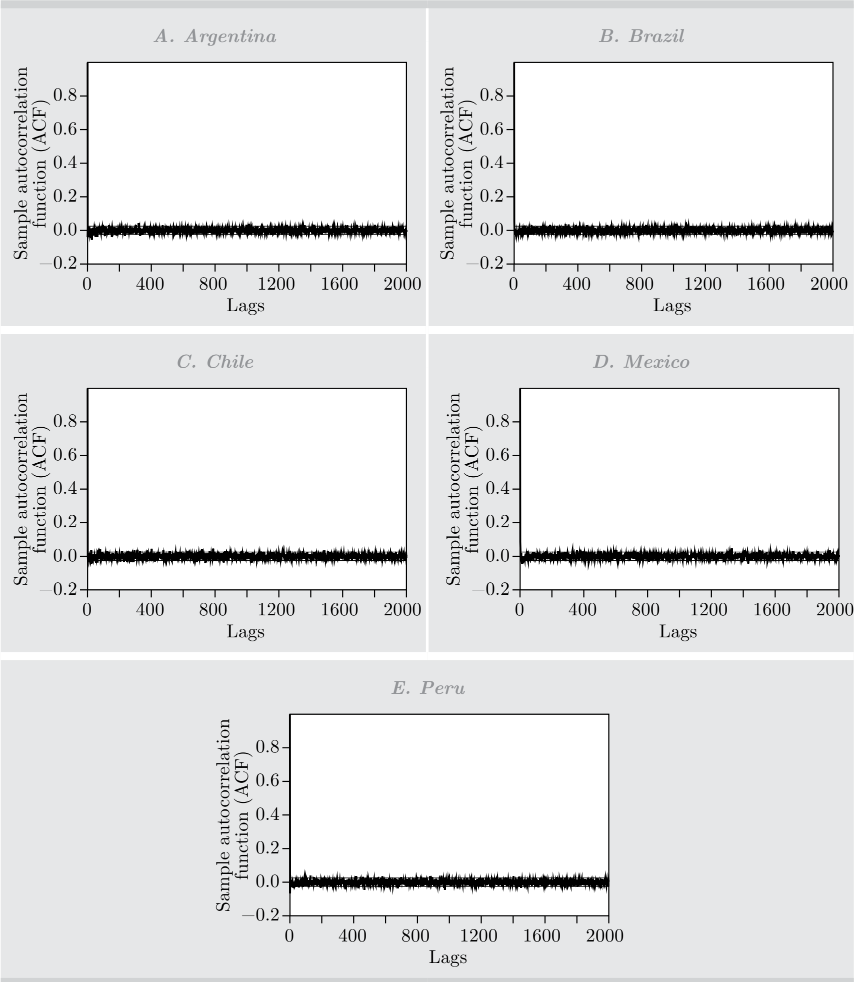
11. The results are available upon request.

Figure 3. Level shift component τ_t estimated by Bai and Perron (2003): solid line; smoothed level shift component: dotted line



Source: Authors' calculations.

Figure 4. Sample ACF of residuals series



Source: Authors' calculations.

in the case of the ARFIMA(1,d,1) model. The volatility series show a positive and significant fractional parameter. Moreover, the parameters ϕ autoregressive and θ moving average are small but significant. In the case of volatility adjusted by the level shift component estimated by Bai and Perron (1998, 2003), the fractional parameter has a negative value that is very close to zero. This confirm that the long-memory behavior is eliminated once the level shifts are taken into account.

3.2 Level shift effect in the GARCH and CGARCH models

Given that GARCH models, as well as ARFIMA models, are frequently used to model volatility, we estimate GARCH(1,1) and CGARCH models. The GARCH model is formulated as:

$$\tilde{r}_t = \sigma_t \epsilon_t, \tag{3}$$

$$\sigma_t^2 = \mu + \beta_1 \tilde{r}_{t-1}^2 + \beta_2 \sigma_{t-1}^2, \tag{4}$$

where $\tilde{r}_t$ are returns discounted by their mean and ϵ_t is a *i.i.d.* t-Student's distribution with mean 0 and variance 1. The CGARCH model is specified as follows:

$$\tilde{r}_t = \sigma_t \epsilon_t, \tag{5}$$

$$(\sigma_t^2 - n_t) = \beta_1(\tilde{r}_{t-1}^2 - n_{t-1}) + \beta_2(\sigma_{t-1}^2 - n_{t-1}), \tag{6}$$

$$n_t = \mu + \rho(n_{t-1} - \mu) + \psi(\tilde{r}_{t-1}^2 - \sigma_{t-1}^2). \tag{7}$$

The important coefficients are β_1 and β_2, which indicates the presence of conditional heteroskedasticity effects. The parameter μ is a constant to which n_t converges, which represents the long-term component of time-varying volatility. Moreover, Equation (6) represents the transitory component of volatility. In addition, the parameter ρ measures the persistence of shocks in the permanent component of Equation (7), while persistence is measured by $(\beta_1 + \beta_2)$ in Equation (4) and the transitory component in Equation (6).

On the other hand, a CGARCH model is estimated but increased with auxiliary (dummy) variables:

$$\tilde{r}_t = \sigma_t \epsilon_t, \tag{8}$$

$$(\sigma_t^2 - n_t) = \beta_1(\tilde{r}_{t-1}^2 - n_{t-1}) + \beta_2(\sigma_{t-1}^2 - n_{t-1}), \tag{9}$$

$$n_t = \mu + \rho(n_{t-1} - \mu) + \psi(\tilde{r}_{t-1}^2 - \sigma_{t-1}^2) + \sum_{i=2}^{m+1} D_{i,t}\gamma_i, \tag{10}$$

where $D_{i,t} = 1$ if t pertains to the regime i, and 0 if it does not pertain to the regime, with $t \in \{T_i + 1, ..., T_{i+1}\}$ and T_i $(i = 1, ..., m)$ with the level shift dates estimated using the Bai and Perron method (1998, 2003). The coefficients γ_i are estimated alongside the parameters of the GARCH model and indicate the size of the level shifts.

The results are presented in Table 3. The parameters β_1 and β_2 in the GARCH model are significant for the five markets analyzed. The parameter β_2 is raised and fluctuates between 0.729 (Peru) and 0.913 (Argentina). The sum of β_1 and β_2 is close to the unit, indicating that the effect of the shocks slowly declines, which is a characteristic frequently found in the estimates. The estimates imply that the half-lives of the shocks are around 77, 29, 99, and 43 days for Brazil, Chile, Mexico and Peru, respectively. In the case of Argentina, the sum of the coefficients is the unity, implying a half-life of the shocks equal to infinity days.

In the CGARCH model we fnd two scenarios for the estimates of β_1 and β_2. For the cases of Argentina, Brazil and Mexico, these coefficients are not significant. On the other hand, for Chile and Peru they are significant but their sum is reduced to around 0.50 (Chile) and 0.80 (Peru). Overall, the half-life of the shocks is reduced. This fact is interesting as it shows that in the CGARCH model, when the volatility is divided into a long-term component and another short-term component, the coefficients β_1 and β_2 are not important, with the so-called GARCH effects disappearing. Nonetheless, the coefficients linked to the long term (ρ, ψ) are highly significant. In particular, it is important to note that the parameter ρ is located very close to the unit. This indicates high persistence in all financial markets analyzed.

Table 3. Estimates of GARCH and CGARCH models

Model	Parameter	Value	S.E.	p-values
Argentina				
GARCH	β_1	0.092	0.005	0.000
	β_2	0.913	0.004	0.000
CGARCH	β_1	-0.002	0.012	0.900
	β_2	-0.516	5.312	0.922
	ρ	1.000	6.180	0.000
	ψ	0.079	0.004	0.000
CGARCH (using τ_t from Bai and Perron)	β_1	-0.018	0.011	0.107
	β_2	-0.671	0.117	0.000
	ρ	0.631	0.036	0.000
	ψ	0.015	0.016	0.000
CGARCH (using smoothed estimate of τ_t)	β_1	0.013	0.012	0.259
	β_2	0.004	0.600	0.995
	ρ	0.770	3.820	0.000
	ψ	0.212	0.009	0.000
Brazil				
GARCH	β_1	0.095	0.008	0.000
	β_2	0.896	0.008	0.000
CGARCH	β_1	-0.026	0.018	0.143
	β_2	-0.179	0.644	0.781
	ρ	0.991	0.004	0.000
	ψ	0.100	0.009	0.000
CGARCH (using τ_t from Bai and Perron)	β_1	0.005	0.028	0.865
	β_2	0.037	3.154	0.991
	ρ	0.543	0.049	0.000
	ψ	0.075	0.026	0.004
CGARCH (using smoothed estimate of τ_t)	β_1	-1.394	7.658	0.856
	β_2	2.170	7.729	0.779
	ρ	0.795	0.036	0.000
	ψ	1.446	7.660	0.852

Table 3. (continued)

Model	Parameter	Value	S.E.	p-values
	Chile			
GARCH	β_1	0.169	0.013	0.000
	β_2	0.807	0.012	0.000
CGARCH	β_1	0.138	0.022	0.000
	β_2	0.359	0.109	0.001
	ρ	0.983	0.005	0.000
	ψ	0.120	0.014	0.000
CGARCH (using τ_t from Bai and Perron)	β_1	-1.095	0.484	0.024
	β_2	1.321	0.551	0.016
	ρ	0.261	0.049	0.000
	ψ	1.235	0.489	0.011
CGARCH (using smoothed estimate of τ_t)	β_1	-0.039	0.406	0.924
	β_2	0.515	1.676	0.759
	ρ	0.625	0.087	0.000
	ψ	0.246	0.401	0.504
	Mexico			
GARCH	β_1	0.084	0.008	0.000
	β_2	0.909	0.008	0.000
CGARCH	β_1	0.016	0.017	0.341
	β_2	-0.650	0.507	0.199
	ρ	0.992	0.004	0.000
	ψ	0.082	0.008	0.000
CGARCH (using τ_t from Bai and Perron)	β_1	0.037	0.041	0.365
	β_2	0.018	0.528	0.974
	ρ	0.515	0.022	0.000
	ψ	0.060	0.038	0.120
CGARCH (using smoothed estimate of τ_t)	β_1	-0.740	1.905	0.698
	β_2	1.484	1.986	0.455
	ρ	0.774	0.006	0.000
	ψ	0.811	1.910	0.671

Table 3. (continued)

Model	Parameter	Value	S.E.	p-values
	Peru			
GARCH	β_1	0.255	0.010	0.000
	β_2	0.729	0.008	0.000
CGARCH	β_1	0.232	0.023	0.000
	β_2	0.583	0.043	0.000
	ρ	0.996	0.004	0.000
	ψ	0.104	0.022	0.000
CGARCH (using τ_t from Bai and Perron)	β_1	0.023	0.093	0.804
	β_2	0.426	1.169	0.716
	ρ	0.730	0.040	0.000
	ψ	0.230	0.094	0.014
CGARCH (using smoothed estimate of τ_t)	β_1	-0.024	0.018	0.171
	β_2	0.024	0.448	0.957
	ρ	0.723	0.000	0.000
	ψ	0.306	0.021	0.000

Source: Authors' calculations.

In fact, the half-lives of the shocks are 77, 41, 87, and 173 days for Brazil, Chile, Mexico and Peru, respectively. In the case of Argentina we again find a half-life of the shocks equal to infinity days.

However, since the level shifts are taken into consideration in the form of dummy variables, the parameters β_1 and β_2 are not significant for any of the series except for the case of Chile. Nonetheless, the sum of both coefficients is much less than the unit. On the other hand, it is evident that the value of the parameter ρ declines drastically from 0.73 (Peru) to 0.26 (Chile). This shows that even when this parameter is significant, the impact of the shocks declines more rapidly than when the level shifts are not considered. The half-lives of the shocks are 1.5, 1.1, 0.52, 1.0, and 2.2 days for Argentina, Brazil, Chile, Mexico and Peru, respectively[12].

In addition, we analyze the sensitivity of the results using the smoothed estimator of the trend component. This is done by replacing the term

12. Using the smoothed estimate of the component of level shifts, the estimates are very similar: 2.7, 3.0, 1.5, 2.7 and 2.1 for Argentina, Brazil, Chile, Mexico and Peru, respectively.

$\Sigma_{i=2}^{m+1} D_{i,t}\gamma_i$, with the smoothed estimator (Gaussian kernel) of the level shift component. The results are similar to those obtained previously. The parameters β_1 and β_2 are not significant (this time including Chile) and the value of the parameter ρ–even if it is significant–decreases drastically, reducing the impact of the permanent component of the equation.

Some conclusions can be ventured thus far: (i) the RLS model with a stationary AR(1) component provides an adequate description of the data; (ii) the level shift component is important and explains both the long memory aspect and the presence of conditional heteroskedasticity as they are generally perceived in the literature. As a final test, we will look at whether the RLS model provides reasonable predictions compared with some traditional models.

4. Forecasting

In this section the RLS model is assessed in comparison with ARFIMA models, with respect to prediction capacity. The predictions are based on the approximation of Lu and Perron (2010). In this way, the τ-periods' forward predictions are given by:

$$\hat{y}_{t+\tau|t} = y_t + HF^{\tau}[\sum_{i=1}^{2}\sum_{j=1}^{2}\Pr(s_{t+1}=j)\Pr(s_t=i\,|\,Y_t)X_{t|t}^{ij}], \tag{11}$$

where $E_t(y_{t+\tau}) = \hat{y}_{t+\tau|t}$ is the prediction of volatility in time $t+\tau$, conditional to information on time t, and the matrices F and H are defined as before and the prediction horizons are $\tau = 1$, 5, 10, 20, 50 and 100. Moreover, as a criterion for measuring prediction confidence, we use the mean squared forecast error (MSFE) proposed by Hansen and Lunde (2006) and defined by:

$$MSFE_{\tau,i} = \frac{1}{T_{out}}\sum_{t=1}^{T_{out}}(\bar{\sigma}_{t,\tau}^2 - \bar{y}_{t+\tau,i|t})^2, \tag{12}$$

where T_{out} is the number of predictions $\bar{\sigma}_{t,\tau}^2 = \Sigma_{s=1}^{\tau} y_{t+s}$, and $\bar{y}_{t+\tau,i|t} = \Sigma_{s=1}^{\tau}\hat{y}_{t+s,i|t}$, with i representing each model. The evaluation and comparison are performed using 5% of the model confidence set

(MCS) proposed by Hansen *et al.* (2011). The MCS allows for better evaluations of the models than can be done with comparisons between pairs of models. One of the advantages of this procedure is that the evaluations are performed by taking into account the limitations of the data. This means that if the data are clear, a single model will be selected; when the data are not sufficiently informative, an MCS with various models would be the result. In these cases we can establish that more than one model offers a good prediction, which cannot be established using other kinds of comparisons.

To perform the predictions, the observations from January 2, 2006 up through the end of the sample were retained. This period includes the international financial crisis and can serve to verify whether the RLS model is a good predictor. The results are presented in Table 4, and lead to the conclusion that the RLS model is within 5% of the MCS for practically all prediction horizons.

Table 4. Comparison of forecasts $\widehat{y}_{t+\tau|t}$

	$\tau = 1$	$\tau = 5$	$\tau = 10$	$\tau = 20$	$\tau = 50$	$\tau = 100$
Argentina						
RLS	0.75	4.73	12.53	37.64	214.88	886.66
	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)
ARFIMA (0,d,0)	0.94	7.32	21.13	65.67	322.63	1096.05
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ARFIMA (1,d,1)	1.23	14.53	50.05	181.17	1025.11	3833.36
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Brazil						
RLS	0.69	3.92	10.04	31.05	175.25	773.83
	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)
ARFIMA (0,d,0)	0.93	8.37	26.76	92.74	498.01	1792.57
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ARFIMA (1,d,1)	0.88	7.22	22.14	74.18	383.24	1341.15
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Chile						
RLS	0.45	4.08	11.91	40.69	260.54	1002.64
	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)
ARFIMA (0,d,0)	0.70	6.31	18.68	58.95	268.86	769.50
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ARFIMA (1,d,1)	0.71	6.29	18.56	58.47	266.00	760.18
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Table 4. (continued)

	$\tau = 1$	$\tau = 5$	$\tau = 10$	$\tau = 20$	$\tau = 50$	$\tau = 100$
Mexico						
RLS	0.57	4.07	11.90	40.19	231.50	886.56
	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)
ARFIMA (0,d,0)	0.79	7.08	22.73	77.58	394.14	1319.17
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ARFIMA (1,d,1)	0.79	7.11	22.83	77.93	395.49	1322.16
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Peru						
RLS	0.49	4.35	11.94	31.94	142.01	620.40
	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)
ARFIMA (0,d,0)	0.80	6.86	19.72	58.01	271.13	927.40
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ARFIMA (1,d,1)	0.80	6.87	19.84	58.91	283.42	1002.74
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Source: Authors' calculations.
Note: Numbers are the MSFE; p-values of the MCS are reported in parentheses; * denotes that the model belongs to the 5% of the MCS of Hansen *et al.* (2011), comparing all models.

5. Conclusions

In this paper, we estimate an RLS model using the approach of Lu and Perron (2010) and Li and Perron (2013) for the volatilities of the financial returns of five Latin American economies. Even though we have fewer observations in comparison with developed countries, our results are conclusive and in line with the findings of Lu and Perron (2010). The estimation results show that the probability of level shifts is small but is responsible for the presence of long memory in the volatilities of the series analyzed. Having estimated the probability of level shifts, the exact number of such level shifts can be calculated. Thus, the component obtained as a difference between the volatility series and the level shifts possesses an ACF that indicates an absence of long memory. We show that short memory processes contaminated with random level shifts may be confused with long memory in the data considered. The estimates of autoregressive conditional heteroskedasticity models discounted by level shifts shows that these components are artificially introduced by level shifts because the estimates of the fractional parameter are negative or close to zero.

Finally, an out-of-sample forecasting exercise shows that the RLS model performs better than traditional long memory models such as the ARFIMA (p,d,q) models.

REFERENCES

Bai, J. and P. Perron (1998), “Estimating and testing linear models with multiple structural changes,” *Econometrica* **66**: 47–78.

Bai J. and P. Perron (2003), “Computation and analysis of multiple structural change models,” *Journal of Applied Econometrics* **18**: 1–22.

Baillie R., T. Bollerslev, and H. Mikkelsen (1996), “Fractionally integrated generalized autoregressive conditional heteroskedasticity,” *Journal of Econometrics* **73**: 3–30.

Bollerslev T. and H. Mikkelsen (1996), “Modeling and pricing long memory in stock market volatility,” *Journal of Econometrics* **73**: 151–84.

Diebold F. and A. Inoue (2001), “Long memory and regime switching,” *Journal of Econometrics* **105**: 131–59.

Ding Z., R. Engle, and C. Granger (1993), “A long memory property of stock market returns and a new model,” *Journal of Empirical Finance* **1**: 83–106.

Engle, R.F. and A.D. Smith (1999), “Stochastic permanent breaks,” *Review of Economics and Statistics* **81**: 553–74.

Fuller, W.A. (1996), *Introduction to Time Series*, 2nd edition, New York: John Wiley.

Geweke J. and S. Porter-Hudak (1983), “The estimation and applications of long memory time series models,” *Journal of Time Series Analysis* **4**: 189–209.

Gourieroux C. and J. Jasiak (2001), “Memory and infrequent breaks,” *Economic Letters* **70**: 29–41.

Granger, C.W.J. and N. Hyung (2004), “Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns,” *Journal of Empirical Finance* **11**: 399–421.

Granger C.W. and R. Joyeux (1980), “An introduction to long memory time series models and fractional differencing,” *Journal of Time Series Analysis* **1**: 15–29.

Hamilton, J. D. (1989), “A new approach to the economic analysis of nonstationary time series and the business cycle,” *Econometrica* **57**: 357–84.

Hamilton, J.D. (1994), *Time series analysis*. Princton, NJ: Princeton University Press.

Hansen, P.R. and A. Lunde (2006), “Consistent ranking of volatility models,” *Journal of Econometrics* **131**, 97–121.

Hansen, P.R., A. Lunde, and J.M. Nason (2011), “The model confidence set, *Econometrica* **79**: 453–97.

Harrison, P. J. and C.F. Stevens (1976), “Bayesian forecasting,” *Journal of the Royal Statistical Society* (Series B), **38**: 205–47.

Hosking J. (1981), “Fraccional differencing,” *Biometrika* **68**: 165–76.

Humala, A. and G. Rodríguez (2013), “Some stylized facts of return in the foreign exchange and stock markets in Peru,” *Studies in Economics and Finance* **30**(2): 139–58.

Li, Y. and P. Perron (2013),"Modeling exchange rate volatility with random level shifts," Working Paper, Boston University.

Lu, Y. and P. Perron (2010), "Modeling and forecasting stock return volatility using a random level shift model," *Journal of Empirical Finance* **17**: 138–56.

Lobato, I. and N. Savin (1998), "Real and spurious long memory properties of stock market data," *Journal of Business and Economics Statistics* **16**: 261–68.

Mikosh, T. and C. Stărică (2004a), "Nonstationarities in financial time series, the long-range dependence and the IGARCH effects," *The Review of Economic and Statistics* **86**: 378–90.

Mikosh, T. and C. Stărică (2004b), "Changes of structure in financial time series and the GARCH model," *Statistical Journal* **2**: 42–73.

Nelson, D. (1991), "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica* **59**(2): 347–70.

Ojeda Cunya, J.A. and G. Rodríguez (2015), "An application of a random level shifts model to the volatility of Peruvian stock and exchange rate returns," forthcoming in *Macroeconomics and Finance in Emerging Market Economies.* It appears as Working Paper 383, Department of Economics, Pontificia Universidad Católica de Peru.

Perron, P. (1989), "The great crash, the oil price shock, and the unit root hypothesis," *Econometrica* **57**: 1361–401.

Perron, P. (1990), "Testing for a unit root in a time series regression with a changing mean. *Journal of Business an Economic Statistics* **8**: 153–62.

Perron, P. and Z. Qu (2010), "Long-memory and level shifts in the volatility of stock market return indices," *Journal of Business and Economic Statistics* **28**: 275–90.

Perron, P. and T. Wada (2009), "Let's take a break: Trends and cycles in U. S. real GDP," *Journal of Monetary Economics* **56:** 749–65.

Robinson, P.M. (1995) "Log-periodogram regression of time series with long range dependence," *The Annals of Statistics* **23**: 1048–73.

Stărică, C. and C. Granger (2005), "Nonstationarities in stock returns," *The Review of Economics and Statistics* **87**: 503–22.

Wada, T. and P. Perron (2006), "An alternative trend-cycle decomposition using a state space model with mixtures of normals: Specifications and applications to international data," Working Paper, Department of Economics, Boston University.

VOL. 52 NO. 2 (NOV, 2015), 213–240

LATIN AMERICAN JOURNAL OF ECONOMICS

ON THE SUSTAINABILITY AND SYNCHRONIZATION OF FISCAL POLICY IN LATIN AMERICA*

PAUL ALAGIDEDE**
GEORGE TWENEBOAH***

This paper explores the sustainability of fiscal policy for a panel of Latin American countries over the period 1990–2012. We extend the literature on the causal relationship between government expenditure (GX) and revenue (GR) in the short run and long run. Our results show a significant long-run relationship between GX and GR, suggesting that fiscal policies are consistent with their intertemporal budget constraints. We establish bidirectional causality between revenue and expenditure in the long run, indicating a contribution from both GX and GR in establishing steady state equilibrium following substantial deviations. Our data also uphold the fiscal synchronization thesis.

JEL classification: C23, E62, H62, H63

Keywords: Fiscal sustainability, Latin America, panel cointegration, fiscal synchronization, intertemporal budget constraints

1. INTRODUCTION

The sustainability of fiscal policy and its implications has received considerable attention in the academic literature and policymaking circles for many years. It is a highly relevant subject because of the role sustainability plays in ensuring financial and macroeconomic stability. Also, a number of financial crisis episodes since the Great Depression of the 1930s have been preceded by rising public debt and fiscal imbalances, notably the debt crisis in Latin America in the early 1980s leading to the so-called "lost decade," and the more recent Eurozone debt crises.

Due to the fundamental relevance of sustainable government spending and restraint to budget deficit financing, the sustainability of fiscal positions has featured in the convergence criteria of a number of monetary and economic pacts. The Maastricht Treaty of 1992 explicitly pegged member governments' public debt and deficit obligation at 60% and 3% of their

* The authors acknowledge the helpful comments provided by anonymous referees.

** Wits Business School, University of the Witwatersrand, 2 St David's Place, Parktown, Johannesburg, South Africa. Email: paul.alagidede@wits.ac.za.

** Wits Business School, University of the Witwatersrand and Department of Banking and Finance, Ghana Baptist University College, Private Mail Bag, Kumasi, Ghana. Email: opokutweneboah@yahoo.com.

doi 10.7764/LAJE.52.2.213

GDP, respectively. The convergence criteria for the establishment of the West African Monetary Zone (WAMZ) also capped the budget deficit at 3% of GDP (see Alagidede *et al.*, 2008 and 2012).

Subsequent to the soaring debt levels of Latin American countries that led to the debt crisis in the 1980s, most governments adopted policies characterized by fiscal rules to guide economic policy. A common feature of the latter part of the 1990s through the early 2000s was the introduction of measures to enhance transparency through a combination of balanced budget targets and/or numerical spending caps. Several authors contend that Latin American countries have generally performed much better in terms of fiscal discipline due to improved fiscal institutional frameworks (Filc and Scartascini, 2007; Eslava, 2012).

In spite of the existence of a plethora of empirical studies on fiscal sustainability in advanced countries and other regions, our knowledge of the subject is far from perfect. Also, given the nature of the debt crisis that hit Latin America in the 1980s and increasing concern about the negative consequences of rising government debt and fiscal deficits, it is imperative that we revisit the question of debt and fiscal sustainability/solvency. The aim of this paper is thus to empirically assess and present lessons on fiscal policy sustainability for a panel of Latin American countries by applying recent advances in the unit root and cointegration literature.

The paper fills a gap by extending the literature on the short-run as well as the long-run causal relationship between government expenditure (GX) and revenue (GR). Using advanced estimation techniques, the relationship is further explored to establish whether countries in Latin America are characterized by either the tax-spend, spend-tax or fiscal synchronization hypothesis, which has critical implications for fiscal sustainability in the region. We employ the Westerlund (2007) panel cointegration tests that impose no common factor restriction, account for possible cross-country dependence, and solve the problems associated with Pedroni's (1999) residual-based tests. Furthermore, a more flexible pooled mean-group (PMG) estimator proposed by Pesaran *et al.* (1999) is specified. This enables us to explore both short-run dynamics and long-run equilibrium relationships among the variables of interest, accounting for non-stationarity in the data and heterogeneity across countries in their short-run dynamic relationships. We compare these with the results obtained using restrictive dynamic fixed effects (DFE) methods, and the more flexible but information-intensive mean-group (MG) approach.

Quite understandably, the causal behavior or relationship between GX and GR may provide practical insights into the dynamics and processes involved in fiscal policy adjustments and how policymakers should approach budget deficits in the future. More importantly, the period under review (1990-2012) captures exactly two decades following the debt crisis. This enables us to assess the effectiveness of the fiscal rules implemented after the crisis and infer how they have shaped the sustainability of the long-run fiscal stance in the region. We report that although GX and GR are non-stationary, they share a common trend. The results show that there is significant causality between the variables in the short run as well as a long-run fiscal synchronization, suggesting that both GX and GR help push the budget towards equilibrium should there be deviations from the long-run relationship.

The remaining sections of this paper are set out as follows: Section 2 reviews the theoretical and empirical literature, while Section 3 presents a description of the data and methodology. In Section 4 the different unit root tests along with the battery of cointegration techniques are explained. The results of the statistical analysis, coupled with the short-run and long-run dynamics of the relationships, are explored. Concluding remarks and policy recommendations are contained in Section 5.

2. Review of empirical literature

2.1. Sustainability of fiscal policy

The ability of a government to sustain its current spending, taxation, and other policies in the long run without threatening default on some of its liabilities or promised expenditures has long occupied economists' attention. A conventional approach applied to establish fiscal policy sustainability has been built around the government's intertemporal budget constraint (IBC) mechanism. If the IBC holds in present value terms, the fiscal policy is considered sustainable. For this to hold, current debt levels must be expected to be compensated by the present value of surpluses garnered from the expected future primary budget. There is a vast literature on this subject but most of the empirical research has focused on the experiences of the United States and other advanced countries (Cuddington, 1997; Chalk and Hemming, 2000), although the conclusion is still not clear (Hakkio and Rush, 1991).

One strand of the literature involves the present value budget constraint (PVBC) approach. The methodology involves testing of the PVBC or the non-Ponzi game (NPG) condition for data on government revenue, expenditure, or fiscal balance. This condition is one of the key assumptions considered within the IBC of the government. Also known as the transversality condition, NPG necessitates that the public debt not grow at a rate greater than the interest rate. If this condition is fulfilled, then the IBC will result in equality between the market value of public debt and the sum of discounted future budget surpluses. If this condition is valid, the theory predicts that the government's fiscal policy will be sustainable. According to Hamilton and Flavin (1986), who pioneered the approach for analyzing the concept of fiscal sustainability, if the present value borrowing constraint is not satisfied, fiscal policy is said to be unsustainable in the long run. Thus, there is sustainability if the PVBC is fulfilled without a significant and sudden shift in the balance of revenue and expenditure to avoid potential liquidity and solvency problems. Most of these studies employ time-series unit root and cointegration analysis to explore whether the present value of IBC is effectively respected. The customary practice in the literature is to examine whether past fiscal balance follows a stationary process or if there is cointegration between government expenditures and revenues (see Hakkio and Rush, 1991 and Trehan and Walsh, 1991).

A number of papers have concentrated on examining the stationarity of fiscal balance (Holmes *et al.*, 2010; Wilcox, 1989; and Hamilton and Flavin, 1986). A stationarity result implies that the sustainability hypothesis holds, whereas a non-stationarity result implies the opposite. Disappointingly, evidence obtained by applying the stationarity approach to fiscal balance has not been found to support the sustainability hypothesis (for example, Vanhorebeek and Rompuy, 1995 and Caporale, 1995). Given that expenditure and revenue exhibit integrated behavior, the second methodology tests for cointegration between these variables (Westerlund and Prohl, 2010; Afonso and Rault, 2010; Ehrhart and Llorca, 2008; Quintos, 1995; Hakkio and Rush, 1991). According to this method, if the series are cointegrated, the sustainability hypothesis is upheld (Prohl and Schneider, 2006; MacDonald, 1992; Haug, 1991).

Recent empirical studies hang on testing for stationarity in the fiscal balance series or cointegration between government expenditure and revenue. Nevertheless, the unit root and cointegration tests used do not normally reject the null of a unit root in the series if there is reason to believe that a country has experienced a structural break in its fiscal policies during the sample period. Additionally, such tests are often said to be of low power in small samples and are suspected of providing poor evidence (Perron, 2006).

The disappointing conclusions from these studies have turned more recent research away from the stationarity approach towards a more flexible econometric test based on cointegration. Under this framework, if government expenditures and revenue are found to be cointegrated with a unit slope coefficient on expenditures, fiscal policy is said to be strongly sustainable. Also, when the slope is less than unity, it is described as being weakly sustainable (Quintos, 1995). Although this attempt has brought some flexibility, the results obtained from this approach have been mixed at best (see, for example, Afonso, 2005; Bravo and Silvestre, 2002; and Papadopoulos *et al.*, 1999).

There have been debates surrounding the causes of failure to establish fiscal sustainability. For their part, Westerlund and Prohl (2007) claim that this failure could be attributed to at least two types of flaws in most previous studies. Since the majority of studies apply techniques designed to test the null of a unit root, they argue that low power in the tests could be one reason why cointegration has been difficult to establish. Again, they contend that most studies employ a country-by-country approach, which doesn't contribute more information to the analysis and essentially disregards the information contained in the cross-sectional dimension. However, they concede that when conventional cointegration tests are applied to each country separately, the results are comparable across countries.

In an attempt to correct these flaws, Westerlund and Prohl (2007) suggest the use of panel unit root and panel cointegration methodologies to generate more precise tests. In the case of the European Union, recent studies based on panel cointegration have provided strong evidence for fiscal sustainability (see Westerlund and Prohl, 2007; Afonso and Rault, 2007; Prohl and Schneider, 2006). Most of these studies have focused on the EU 15 and some have properly accounted for the existence of structural breaks.

There is also some evidence relating to member countries of the Organization for Economic Cooperation and Development (OECD). A study by Ehrhart and Llorca (2007) applied panel cointegration to assess fiscal policy sustainability in a sample of 20 OECD countries. They report that expenditure and revenue are co-integrated, implying consistency in fiscal policies with the intertemporal budget constraint for 1975 to 2005. Again, using quarterly data that covers eight wealthy

OECD countries from 1977 to 2005, they applied panel techniques to establish that the fiscal sustainability hypothesis could not be rejected.

Other regions have also benefited from recent advances in the literature. For example, Adedeji and Thornton (2010) and Lau and Baharumshah (2005) consider Asian countries. These studies have found that although fiscal sustainability could be established for the region, the evidence indicates that such sustainability is "weak" and the authors suggest implementation of policy measures to create a more sustainable basis for public finances. For the Southern Mediterranean region, Ehrhart and Llorca (2006) use recent econometric methodology for panel data to test whether there is long-run sustainability in the fiscal policies in six countries–Egypt, Israel, Lebanon, Morocco, Tunisia and Turkey–establishing that fiscal policies are sustainable in these countries.

2.2. Causal relationships between expenditure and revenue

Another dimension of the empirical literature has focused on the causal relationship between government expenditure and revenue through four different theoretical propositions. If no cointegration is detected, we say that there is no evidence of causality between the variables, implying spending and revenue are not related in the long run. However, if cointegration is established, three different outcomes are possible since causality implies that a change in one variable necessitates or drives a change in the other variable (Engle and Granger, 1987). We can assess whether causality runs from revenue to expenditure, from expenditure to revenue, or in both directions. The tax-spend hypothesis is based on evidence of a unidirectional causality running from revenue to expenditure as championed by Friedman (1978). Friedman argues that tax cuts lead to higher deficits, and if a government cares about the implications of this, it will reduce its level of spending to equal the level of tax revenue or possibly lower.

An alternative version of this hypothesis was advanced by Wagner (1976) and Buchanan and Wagner (1978). Contrary to Friedman (1978), they find that taxes unidirectionally induce negative changes in expenditure. This means that increased taxes would lead to spending cuts. The thrust of the Buchanan and Wagner (1978) argument is that taxpayers suffer from fiscal illusion. The authors point out that when taxes are cut, the taxpayer will assume that the cost of providing goods and services has fallen, and will therefore demand more government programs. If such programs are undertaken, this will result in an increase in government spending. So, while tax changes induce changes in spending, the relationship is an inverse one as postulated by Buchanan and Wagner (1978); this hypothesis prescribes increased taxes as the cure for budget deficits.

The spend-tax hypothesis advanced by Peacock and Wiseman (1979) and Barro (1979) is based on causality directed from expenditure to revenue. Here, the fiscal illusion problem does not apply and proponents argue that an increase in government spending induces tax hikes. On this basis, they suggest that spending cuts are the solution to budget deficit problems. Yet another hypothesis, termed fiscal synchronization, based on Musgrave's (1966) classical view of public finance, argues that there is a bidirectional causal relationship between revenue and expenditure. Under this theory, revenue and expenditure are determined simultaneously and the public is said to understand the benefits of government services in relation to their costs (Musgrave, 1966). The implication of this theory is that the best strategy for dealing with problems of fiscal deficit is to cut spending and undertake intensive measures to increase revenues.

The empirical evidence on this aspect is mixed; studies based on the United States have provided results that are open to debate. While some researchers provide support for the tax-spend hypothesis (Hoover and Shefrin, 1992; Bohn, 1991; Ram, 1988; Blackley, 1986), others have reported findings that sustain the spend-tax hypothesis (Ross and Payne, 1998; Jones and Joulfain, 1991; Anderson *et al.*, 1986). Interestingly, while Owoye (1995), Miller and Russek (1990) and Manage and Marlow (1989) suggest that the fiscal synchronization hypothesis holds, Baghestani and McNown (1994) find no causal association between the variables.

The case of Latin American countries has not been different. Ewing and Payne (1998) find evidence of the fiscal synchronization hypothesis for Chile and Paraguay and report findings of causality from revenue to expenditure for Colombia, Ecuador, and Guatemala. Baffes and Shah (1990, 1994) find similar results of strong bidirectional causality for Brazil and Mexico, while for Chile and Argentina support was identified for causality from expenditure to revenue. A study of eight countries in Latin America by Cheng (1999) reports on feedback causality for Brazil, Chile, Panama, and Peru to suggest that expenditure and revenue are jointly determined. The same study, however, found causality from revenue to expenditure in some countries–Colombia, the Dominican Republic, Honduras, and Paraguay. This is evidence that the question is empirically unresolved.

Although an extensive theoretical and empirical literature has surfaced on the topic in recent years, not much has focused on Latin American countries. There is a large body of academic writing on this subject in Latin America exploring the stabilization programs and political or institutional factors affecting the region's fiscal performance. Interestingly, little work exists on the sustainability of fiscal policies in the region from the panel econometric point of view.

This article follows recent advances in the application of econometrics to fiscal sustainability, employing recently developed linear panel unit root and cointegration techniques to analyze data on government expenditure and revenue for Latin American countries. In order to overcome problems caused by small sample size, we make use of alternative long-run panel estimation techniques.

3. Data and Methodology

Annual data on government expenditure (GX) and revenue (GR) as a percentage of gross domestic product (GDP) are extracted from the World Development Indicators (WDI) database published by the World Bank (2015). The publisher indicates the source organizations as the International Monetary Fund, Government Finance Statistics Yearbook and data files, and World Bank and OECD GDP estimates. The data for government revenue exclude grants (percentage of GDP). Revenue consists of cash receipts from taxes, social contributions, and other revenue such as fines, fees, rent, and income from property or sales. Expenses consist of cash payments for the government's operating activities in providing goods and services. It includes compensation of employees (such as wages and salaries), interest and subsidies, grants, social benefits, and other expenses such as rent and dividends. The available data enable the construction of a balanced panel for six countries in Latin America and the Caribbean–the Bahamas, Brazil, Guatemala, Nicaragua, Peru, and Uruguay–for the period 1990–2012. We do not conduct the analysis on a country basis given the relatively short span of the sample. Also, given the strong links among economies in the region, a panel approach is more appropriate in this context. Some countries in the region have been excluded due to lack of consistent data for a balanced panel structure. Figures 1 and

2 are the graphs of revenue/GDP and expenditure/GDP, respectively, for the countries under consideration.

The countries in the panel have generally recorded negative cash balances as a ratio of GDP. Since 2001, only Nicaragua and Peru have had a few instances of positive fiscal balances, defined as revenue (including grants) minus expenditure and net acquisition of nonfinancial assets. From a record deficit of 34.24% in 1990, Nicaragua recorded the highest surplus of 6.1% in 1991, when Brazil, Guatemala, and Uruguay also recorded surplus cash balances. Subsequently, the performance of these economies has not been encouraging, since most of them have experienced negative budget balances.

Although there have been record fiscal deficits and high public debt levels in Latin America, there were markedly favorable conditions during the period 2003–2007. This resulted from an unusual combination of a financial boom, exceptionally high commodity prices, and strong remittances from migrant workers. Since the year 2002, there has been a general upward trend in revenue with the exception of the general and marked dip in 2008–2009. Interestingly, a number of the region's economies were already experiencing a substantial slowdown over the course of 2008, with only Peru recording a surplus balance of 2.1%. However, in 2009 all of the countries experienced a sharp dip in revenue, which was outstripped by higher expenditure and led to deficit balances. To a great extent, the dip can be attributed to the 2008–2009 financial crisis, which gave rise to a general upswing in government spending levels in 2009. Peru's performance has been very impressive in recent years but the country recorded a deficit of approximately 1.1% in 2009 due to the financial crisis.

The crisis of 2009 affected all of the economies to the extent that primary surpluses declined significantly and pushed up the ratio of outstanding public debt to GDP. Ocampo (2009) argues that the crisis manifested in complex ways over time and had different effects on the different countries in the region. According to the author, the initial impact came around the third quarter of 2007, and consisted of a large decline in capital flows and bond issues, a modest increase in financing costs, and a similarly moderate decline in stock market values. However, most of the economies had recovered to pre-crisis revenue levels by 2010. All of the economies under consideration have seen an upward trend in their revenue, with Brazil, Nicaragua, Peru, and Uruguay recording one of the highest revenue/GDP ratios in 2012 since 1990.

Figure 1. Revenue/GDP of individual countries (in logs)

Source: Based on raw figures from World Development Indicators (online version).

Figure 2. Expenditure/GDP of individual countries (in logs)

Source: Based on raw figures from World Development Indicators (online version).

3.1. Panel unit root and stationarity tests

This section involves the application of a battery of panel unit root and stationarity tests to analyze the properties of the data generation process and verify whether the properties are integrated. Five distinct panel unit root techniques are employed: LLC (Levin *et al.*, 2002), Breitung, IPS (Im *et al.*, 2003), ADF-Fisher, and PP-Fisher (Maddala and Wu, 1999). These tests have been proposed based on different sets of assumptions. Each of these tests has a null hypothesis of unit root. The LLC and Breitung tests are based on a common unit root process hypothesis that the autocorrelation coefficients of the variables are homogeneous across cross sections. On the other hand, the IPS, PP-Fisher and ADF-Fisher techniques are based on the assumption that the autocorrelation coefficients across the sections are heterogeneous. To minimize problems arising from cross-sectional dependence, the cross-sectional means are subtracted in the LLC, IPS and Maddala and Wu tests. The Breitung test allows for cross-sectional dependence. In terms of the country-specific maximum number of lags used for the ADF regressions with respect to the LLC, Breitung and IPS tests, this is determined by the Schwarz-Bayesian information criterion. Also, the long-run variance for the LLC and the maximum lags are determined using the Bartlett kernel and Newey-West bandwidth selection algorithm, respectively.

In addition to the unit root tests, one panel stationarity test proposed by Hadri (2000) is employed. According to Baltagi (2008), the residual-based Lagrange multiplier (LM) test is in fact a panel generalization of the KPSS test proposed by Kwiatkowski *et al.* (1992) for time series data. Maddala and Wu (1999) highlight that the ADF regression tests are sensitive to the choice of lag lengths. Furthermore, both tests assume cross-section independence and therefore constrain the associated AR coefficient so that it is homogeneous across sections. If this strong assumption of cross-sectional independence fails, the results of the tests become misleading. Therefore, the Hadri test uses residuals from individual OLS regressions on deterministic components to compute the LM statistic. The null hypothesis is that the panel data is stationary (i.e., no unit root in any of the time series), versus the alternative of non-stationarity for at least some cross-sections. The test can also allow for a general form of dependence over time and for the disturbance component to be heteroskedastic across individual sections. Table 1 gives a summary of other characteristics of the tests.

Table 1. Properties of panel unit root and stationarity tests

	ADF-Fisher	PP-Fisher	IPS	Breitung	Hadri	LLC
Type of test	Individual	Individual	Individual	Common	Common	Common
H0 hypothesis	UR	UR	UR	UR	Stationary	UR
H1 hypothesis	No UR in some CS	No UR in some CS	No UR in some CS	No UR	UR in some CS	No UR
Assumptions	N, F, T	N, F, T	F, T	T	F, T	N, F, T
Autocorrelation correction	Lags/kernel	Lags/kernel	Lags	Lags	Kernel	Lags
Cross-section dependence	Demean	Demean	Demean	Robust	Robust	Demean
Unbalanced panel	Yes	Yes	Yes	No	No	No

Source: Compiled from Eviews Users Guide (Version 8.1) QMS (2014:510-511).
Notes: N = number of exogenous variables; F = individual/fixed effects; T = individual time effects; individual linear trends; CS = Cross-sections; UR = unit root

3.2. Panel cointegration tests

This study tests for cointegration between government expenditure and revenue in the panel of Latin American countries. Three different methodologies composed of tests with different assumptions are employed. Two of these tests–Pedroni (1999, 2004) and Kao (1999)–are based on the two-step cointegration approach of Engle and Granger (1987) for estimating cointegration of heterogeneous panels. Pedroni uses the residuals from the long-run regression to construct four panel cointegration test statistics that assume homogeneity of the autoregressive (AR) term ("panel statistic" or within-dimension tests) and three panel cointegration test statistics that allow for heterogeneity of the AR term ("group statistics" or between-dimension tests). The panel v-statistic and the panel rho-statistic are comparable to the long-run variance ratio statistic for time series and the semi-parametric rho statistic of Phillips and Perron (1988), respectively. The other two–panel PP-statistic and panel ADF-statistic–are extensions of the non-parametric Phillips-Perron and parametric ADF t-statistics, respectively. The tests are valid for only I(1) variables. They also allow for heterogeneous slope coefficients, fixed effects, and individual specific deterministic trends. The critical values for the null hypothesis of no cointegration are derived by Pedroni (1999).

The Kao test also includes residual-based DF and ADF tests similar to Pedroni's seven tests. However, Kao (1999) specifies the initial regression with individual fixed effects, no deterministic trend, and homogeneous regression coefficients. Although both the Pedroni and Kao tests assume the presence of a single cointegrating vector, the Pedroni tests assume heterogeneity of the vector across individual sections (i.e., countries).

Finally, this study employs the structural panel cointegration methodology developed by Westerlund (2007). The four tests proposed are an extension of Banerjee *et al.* (1998) that allow for heterogeneity in a cointegrating vector for I(1). Westerlund's ECM panel cointegration does not impose any common parameter constraint, unlike the residual-based tests. According to the alternative hypothesis one can distinguish between group-mean tests (Gt and Ga) and panel tests (Pt and Pa).

4. Empirical Results

4.1. Panel unit root testing

Since panel cointegration methodologies assume panel data to be integrated of order 1, we analyze the data generating process (dgp) to ascertain the stationarity properties using the LLC, Breitung, IPS, Hadri, ADF- and PP-Fisher tests. A rejection of the null hypothesis of unit root indicates a stationary process whereas a rejection of the null of stationarity under the Hadri test would indicate presence of unit root. Table 2 shows the results of all the tests, which provide evidence that we cannot reject the hypothesis of unit root processes in both the GX and GR variables for the panel of seven Latin American countries. In addition, the Hadri tests strongly reject the null hypothesis of stationarity. This provides strong evidence that the variables have unit roots (i.e., they are integrated processes).

Table 2. Panel unit root tests for Latin America

	Tests assuming individual unit root process			Tests assuming common unit root process		
	IPS w-stat	**ADF-Fisher χ^2**	**PP-Fisher χ^2**	**LLC t^*-stat**	**Breitung t-stat**	**Hadri z-stat**
GX	**-1.00** [0.16]	13.57 [0.33]	44.51* [0.00]	7.64 [1.00]	-0.77 [0.22]	4.01* [0.00]
GR	0.14 [0.56]	5.91 [0.92]	12.68 [0.39]	5.41 [1.00]	-0.53 [0.30]	3.82* [0.00]
ΔGX	-8.66 [0.00]	74.33 [0.00]	329.55 [0.00]	-6.44 [0.00]	-3.98 [0.00]	4.16 [0.00]
ΔGR	-3.97 [0.00]	35.68 [0.00]	48.44 [0.00]	-3.86 [0.00]	-5.01 [0.00]	3.22 [0.00]

Source: Authors' calculations.
Notes: Max lag for LLC, Breitung and IPS is 4. *, ** and *** represent significance at the 1%, 5% and 10% levels, respectively. Values in parenthesis denote p-values. Δ represents first difference of the variables.

The presence of unit root in GX and GR series indicates that the variables are not stationary in levels. Further tests to confirm the order of integration indicate that the variables are difference-stationary. This random walk behavior implies that revenue and expenditure grow without bounds over time and that random shocks to the data-generating process have a permanent effect on the variables. Some have argued that because fiscal sustainability requires that government expenditure and revenue are integrated of order zero, it can be said that fiscal policies do not satisfy the IBC conditionality and for that matter, the strong form of fiscal sustainability would not hold (Shinnick, 2008). However, strictly speaking what is required is that the fiscal balance be stationary so that public debt does not grow beyond the repayment limit, which can be achieved as long as the debt is stationary. This in turn would indicate that all that is needed for sustainability is that revenue and expenditure cointegrate (Munawar-Shah and Abdul-Majid, 2014).

4.2. Panel cointegration testing

After establishing the data-generation process of the variables, we proceed to test whether the logarithm of revenue (GR) and its covariates as well as the logarithm of expenditure (GX) and its associated covariates share a common stochastic trend. Three alternative panel cointegration techniques are employed for this purpose. They include two tests based on the residuals of the long-run static regression (Pedroni and Kao) and the Westerlund ECM panel cointegration tests. The Bayesian information criterion is used to automatically select the appropriate lag length for Pedroni and Kao tests. We include deterministic time trends in all specifications and select the Bartlett kernel bandwidth with the Newey-West algorithm.

The results in Table 3 provide strong support for the presence of cointegration when both GR and GX are used as the dependent variable, at least at the 5% significance level. This evidence further indicates the possibility of a somewhat bi-directional long-run equilibrium

Table 3. Pedroni and Kao residual cointegration tests

(Dependent variable: GR)

Test	Panel test statistics	Prob.	Test	Group test statistics	Prob.
Panel v	3.08**	[0.01]	Group rho	-3.62*	[0.00]
Panel rho	-5.33*	[0.00]	Group PP	-6.03*	[0.00]
Panel PP	-6.18*	[0.00]	Group ADF	-7.01*	[0.00]
Panel ADF	-6.63*	[0.00]	Kao	-9.86*	[0.00]

Source: Authors' calculations.
Notes: Test results were generated by Eviews 8. Pedroni's panel statistics are weighted. Values in [ ] are *p*-values. *, **, and *** indicate significance at 1%, 5% and 10%, respectively.

Table 4. Westerlund ECM panel cointegration tests

Statistic	Dependent variable: GR				Dependent variable: GX		
	Value	Z-value	P-value		Value	Z-value	P-value
Gt	-1.74	-1.80**	[0.04]	Gt	-2.09	-2.62*	[0.00]
Ga	-5.33	-0.82	[0.21]	Ga	-8.05	-2.29**	[0.01]
Pt	-4.68	-2.94*	[0.00]	Pt	-4.43	-2.72*	[0.00]
Pa	-6.27	-4.43*	[0.00]	Pa	-8.14	-6.02*	[0.00]

Source: Authors' calculations.
Note: Results generated by Stata 12. *P*-values are in parenthesis. *, ** and *** indicate significance at 1%, 5% and 10% significance level, respectively.

relationship between revenue and expenditure. Table 4 reports the results of the Westerlund tests, which take into account cross-sectional dependencies. It also provides evidence of cointegration, suggesting that fiscal policies in the region for the period under review are sustainable.

4.3. Panel cointegration estimation

The study proceeds to estimate the short-run and long-run coefficients to investigate the causal relationship between GR and GX after establishing the existence of a cointegration relationship between the variables. We also address possible reverse causality between the two variables. In order to ensure a robust analysis, the results of four alternative estimation strategies are reported–the dynamic OLS (DOLS), mean group, pooled mean group, and dynamic fixed effects. Saikkonen (1991) and Stock and Watson (1993) originally proposed the DOLS estimator, which was later generalized by Kao and Chiang (2000). The estimation involves augmenting a static long-run relation by leads and lags of first-differenced explanatory variables. This strategy improves the efficiency of the long-run estimates, although it does not capture the short-run dynamics. Therefore, we include the PMG estimator proposed by Pesaran *et al.* (1999). The estimator is a panel extension of the single equation autoregressive distributed lag (ARDL) model, which has the advantage of the error correction representation. It provides information about the contemporaneous impacts and the speed of adjustment towards the long-run equilibrium state after a disturbance. Furthermore, while the long-run coefficients are assumed to be identical across panels (homogeneous), the short-run coefficients are allowed to vary across the sections of the panel (heterogeneous) (see Bangake and Eggoh, 2012). Also, the MG estimator which allows the long-run parameters to be heterogeneous is employed. The DFE estimator which assumes homogeneity for both the short- and long-run parameters is included.

4.4. Comparing the results of PMG, MG and DFE estimates

Since the cointegration estimators have different assumptions and impose different restrictions there is an important test used to measure and compare the efficiency and consistency of the PMG, MG, and DFE estimations. Another reason to do this is that the estimators usually report different and sometimes contradictory results along with the restrictions. Given our results for the GX model in Table 5, we realize that although the signs (or directions) of causality are consistent among the three estimators, the magnitude (or size) as well as the degrees of significance differ slightly. For instance, whereas both the MG and DFE give a convergence coefficient of more than 0.60, the PMG estimator reports that it is 0.51. Also, we find that whereas the PMG and DFE estimators show much higher long-run coefficients of GX–0.97 and 0.94, respectively–the MG produces a lower coefficient of 0.86. Again, when the results for the GR model are compared, the DFE estimates are quite similar to the PMG results. In the long run, both PMG and DFE produce coefficients of around 0.70 whereas the MG gives a GR coefficient of 0.21. On the basis of this finding, it becomes imperative that we choose the one that is more efficient and consistent for the analysis.

We apply the Hausman h-test to examine the efficiency of the PMG estimator compared to the other estimators and the validity of the long-run homogeneity restriction across countries. The test has a null hypothesis that the difference between the PMG and MG estimation or the PMG and DFE estimation is not systematic. A failure to reject the null indicates that the PMG estimator is recommended, as it is more efficient under the null hypothesis. If the alternative applies–that there is a significant difference between PMG and MG or PMG and DFE–the null is rejected. If the results indicate that the p-value is insignificant at the 5% level, then the PMG will be used. However, if the p-value becomes significant, then the use of MG or DFE estimator is deemed appropriate.

In the GX model, the test indicates that the PMG estimator is favored since the null hypothesis cannot be rejected at the 1% significance level. Also, between PMG and DFE, the PMG is favored. Hence, our analysis for this model is based on the PMG estimations. The test further indicates that the PMG is the preferred estimator among the other GR specifications, so it is employed for our subsequent analysis.

4.5. Short-run causality and convergence dynamics

The results indicate that the lag of the expenditure variable has a positive impact on the current values of revenue. This means that an increase in expenditure causes a hike in revenue. Similarly, the lag of the revenue variable has a positive impact on the current values of expenditure; we find that an increase in revenue causes a rise in expenditure. In both cases the *p*-values indicate that the coefficients are significant. This means that the effect of either expenditure or revenue on the other variable is statistically significant in the short run, which suggests strong evidence to support the claim that there is short-run causality between GX and GR.

In all three instances, the error correction terms or convergence coefficients that capture the speed of adjustment are statistically significant at the 1% level. This strong significance lends more support to the evidence of a long-run relationship or causality between the variables. This means further evidence of cointegration is established by the error correction term (convergence coefficient), which is statistically significant. The error correction terms are negative, which is expected as it implies that, for any deviations of expenditure in the previous period from the long-run equilibrium, the error correction term stimulates a positive change in revenue to revert back to the original equilibrium. In the same manner, if revenue in the past period overshoots the equilibrium, then it is forced to come back towards equilibrium. Also, the somewhat large magnitudes imply that the model returns to its equilibrium state quickly after an unexpected shock or deviation; both GX and GR adjust in response to deviations and approach the long-run equilibrium condition. This has actually been the case since most Latin American countries implemented fundamental fiscal-institutional reforms and adopted fiscal frameworks in the form of numerical rules that placed constraints on debt, deficits and/or expenditure and procedural rules and transparency rules aimed at establishing fiscal consolidation and budgetary discipline (see Hallerberg and Scartascini (2011) for details). There have been claims that the significant progress made in fiscal discipline in the mid-1980s, as pointed out by Edwards (1995), has had important positive consequences for Latin America (Sanchez, 2003). The main thrust of fiscal adjustment took place in the first half of the 1990s (Sanchez, 2003). In essence, since this study covers the period during which most of the reforms were implemented, we posit that the institutional and fiscal policy reforms in the region may have been effective in ensuring the high speed of adjustment towards fiscal sustainability.

4.6. Long-run fiscal synchronization

Table 5 indicates that the long-run coefficients are positive and statistically significant, which indicates that GR and GX have a significant positive impact on each other and an increase in GR or GX would bring about a response from the other variable in a similar direction. This supports the evidence of long-run fiscal synchronization hypothesis. The fiscal synchronization hypothesis asserts that expenditure and revenue decisions are made simultaneously by national authorities. It implies that, in an attempt to tackle the problem associated with persistent, rising levels of budget deficit, Latin American governments need to be cautious, as pointed out by Manage and Marlow (1986), about simply cutting expenditures, increasing revenue, or simply altering both revenues and expenditures without taking into consideration the dependence of one variable on the other.

Our evidence lends support to similar studies such as Owoye (1995), Bhat *et al.* (1993), Manage and Marlow (1986), Joulfaian and Mookerjee (1990), and Nyamongo *et al.* (2007). For Latin American countries, the finding is in line with Ewing and Payne (1998), Baffes and Shah (1990, 1994), and Cheng (1999) who provide evidence for feedback causality between expenditure and revenue in support of the fiscal synchronization hypothesis.

4.7. Evidence of weak-form sustainability in the long run

According to Quintos (1995), there is a difference between strong sustainability and a weak form of fiscal sustainability. Hence, we estimate the coefficient of the long-run relation between GR and GX. A strong solvency occurs if there is cointegration and the slope coefficient β of GX is unity. Also, a weak solvency is confirmed when β is less than unity. In this context only the strong condition is appropriate to assess fiscal sustainability (Hakkio and Rush, 1991). This is because the weak condition may be satisfied even as the governments face challenges

Table 5. Panel cointegration estimation results

Dependent variable		GX		
	DOLS	MG	PMG	DFE
Convergence coefficients	N/A	-0.68* (0.00)	-0.51* (0.00)	-0.65* (0.00)
Long-run coefficients	0.99* (0.06)	0.86* (0.15)	0.97* (0.08)	0.94* (0.07)
Short-run coefficients	N/A	0.16 (0.17)	0.25** (0.11)	0.24* (0.09)
Dependent variable		**GR**		
Convergence coefficients	N/A	-0.45* (0.00)	-0.46* (0.00)	-0.47* (0.00)
Long-run coefficients	0.85* (0.05)	0.21 (0.53)	0.73* (0.05)	0.75* (0.07)
Short-run coefficients	N/A	0.27** (0.10)	0.24** (0.11)	0.28* (0.05)
Hausman test	GX	GR		
MG vs. PMG	0.86 (0.36)	0.61 (0.44)		
PMG vs. DFE	0.52 (0.47)	0.23 (0.63)		

Source: Authors' calculations.
Notes: Values in ( ) are standard errors. The xtpmg command in Stata 12 is used for MG, PMG and DFE estimators. *, **, and *** indicate significance at the 1%, 5% and 10% levels, respectively. For the Hausman test and convergence coefficients, values in ( ) are p-values for [ec] and X2, respectively.

financing fiscal deficits, if the revenue relative to GDP is continuously exceeded by expenditure as a percentage of GDP.

In order to achieve this objective, we also test whether the coefficient of GX in the GR model is significantly different from 0. The long-run coefficients are reported in Table 5. From the table, it can be said that the estimated β of GX is 0.73, which is not too far from unity. Further tests on the model reject both the null hypothesis of $\beta = 0$ and that of $\beta = 1$ at the conventional significance levels. Hence, while two non-stationary variables, GR and GX, are cointegrated in the panel of Latin American economies, they can best be judged to be sustainable only in the weak form. We argue that, on the basis of causality, a rise in GR causes a rise in GX and vice versa. However, the magnitude of changes in GR and GX differ. From our analysis, a 1% increase in GX causes GR to increase by less than 1%, which implies that although sustainable fiscal positions are feasible, governments in the region spend more than they receive in revenue. By extension, the results imply that if governments spend at a lower rate compared to their ability to raise revenue in the long run, so that GX and GR are one-to-one, then the strong-form sustainability can be confirmed and there would be no cause for alarm about the future course of a fiscal deficit situation.

5. Concluding remarks and policy recommendations

This study is a contribution to the empirical literature on fiscal sustainability. We make use of recent advances in time series econometric techniques to test whether fiscal policies executed in Latin America over the period 1990–2012 are sustainable in the long run. Tests for panel data for the Bahamas, Brazil, Guatemala, Nicaragua, Peru, and Uruguay in the form of unit roots and cointegration were applied and the results indicate that while both government expenditure (GX) and government revenue (GR) contain unit roots, they have a significant relationship in the long run. This means that fiscal policies in the region are in harmony with their intertemporal budget constraints, indicating the ability to repay financial obligations in the form of debt without explicit default. Sustainable fiscal policies can be continued without changes in policy directions, particularly when there is validity of intertemporal budget constraint in present value terms. However, this long-run sustainability is only in the weak form.

The results show that there is significant causality between expenditure and revenue in the short run as well as long-run bidirectional causality between them, suggesting that both GX and GR help push the budget towards equilibrium in the event of deviations from the long-run relationship. This finding supports the hypothesis of fiscal synchronization, demonstrating the impact fiscal and institutional reforms have had on budgetary outcomes in the region over the study period. To be able to tackle the issue of persistent fiscal deficits in the region, policymakers need to devise strategies to increase revenue and moderate government spending concurrently, as the results point to weak-form sustainability. Consistent with the common caveat in panel cointegration literature, we note that our results are not to be taken out of the regional context to suggest that individual countries within Latin America have pursued sustainable fiscal policies. The sustainability of individual countries may not be achieved if the government's past fiscal behavior remains unbalanced in the long run.

This study enjoys the advantage of the panel approach and points to the solvency of fiscal policies, providing relevant, practical insights into the dynamics of fiscal policy adjustments in the absence of a common fiscal policy in the region. We uphold that future studies should consider models consistent with independent national fiscal policies whenever the available data allow for such analysis.

REFERENCES

Adedeji, O. and J. Thornton (2010), "Fiscal sustainability in a panel of Asian countries," *Applied Economics Letters* **17**: 711–15.

Afonso, A. (2005), "Fiscal sustainability: the unpleasant European case," *Finanz-archiv* **61**: 19–44.

Afonso, A. and C. Rault (2010), "What do we really know about fiscal sustainability in the EU? A panel data diagnostic," *Review of World Economics* **145**:731–55.

Alagidede, P., S. Coleman and J.C. Cuetas (2012), "Persistence of inflationary shocks and common economic trends: Implications for West African monetary union membership," *Journal of Policy Modeling* **34**: 460–75.

Alagidede, P., G. Tweneboah and A. M. Adam (2008), "Nominal exchange rates and price convergence in the West African monetary zone," *International Journal of Business and Economics* **7**: 181–98.

Anderson, W., M.S. Wallace and J.T. Warner (1986), "Government spending and taxation: What causes what?" *Southern Economic Journal* **52**: 630–39.

Asteriou, S. and G. Hall (2007), *Applied econometrics: A modern approach using Eviews and Microfit.* London: Palgrave Macmillan.

Baffes, J. and A. Shah (1990), "Taxing choice in deficit reduction," World Bank Working Paper Series No. 556. Washington, D. C.: The World Bank.

Baffes, J. and A. Shah (1994), "Causality and co movement between taxes and expenditures: Historical evidence from Argentina, Brazil and Mexico," *Journal of Development Economics* **44**: 311–31.

Baghestani, H. and R. McNown (1994), "Do revenues or expenditures respond to budget disequilibria?" *Southern Economic Journal* **61**: 311–22.

Baltagi, B.H. (2008), *Econometric analysis of panel data.* Chichester: John Wiley & Sons.

Baltagi, B.H., J.M. Griffin and W. Xiong (2000), "To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand," *The Review of Economics and Statistics* **82**:117–26.

Banerjee, A., J.J. Dolado and R. Mestre (1998), "Error-correction mechanism tests for cointegration in a single-equation framework," *Journal of Time Series Analysis* **19**: 267–83.

Bangake, C. and J.C. Eggoh (2012), "Pooled mean group estimation on international capital mobility in African countries," *Research in Economics* **66**: 7–17.

Barro, R.J. (1979), "On the determination of public debt," *Journal of Political Economy* **81:** 940–71.

Bhat, K.S., V. Nirmala and B. Kamaiah (1993), "Causality between tax revenue and expenditure of Indian states," *The Indian Economic Journal* **40**: 109–17.

Blackley, R.P. (1986), "Causality between revenues and expenditures and the size of the federal budget," *Public Finance Quarterly* **14**: 139–56.

Bohn, H. (1991), "Budget balance through revenue or spending adjustments: Some historical evidence for the United States," *Journal of Monetary Economics* 27: 333–59.

Bravo, A. and A. Silvestre (2002), "Intertemporal sustainability of fiscal policies: Some tests for European countries," *European Journal of Political Economy* **18**: 517–28.

Breitung, J. (2000), "The local power of some unit root tests for panel data," in B. Baltagi, ed., *Nonstationary panels, panel cointegration, and dynamic panels*: *Advances in econometrics*, V.15. Amsterdam: John Adams Institute.

Buchanan, J.M. and R.W. Wagner (1978), "Dialogues concerning fiscal religion," *Journal of Monetary Economics* **4**: 627–36

Caporale, G. (1995), "Bubble finance and debt sustainability: A test of the government's intertemporal budget constraint," *Applied Economics* **27**: 1135–43.

Chalk, A. and R. Hemming (2000), "Assessing fiscal sustainability in theory and practice," IMF Working Paper 00/81. Washington, D. C.: International Monetary Fund.

Cheng, B.S. (1999), "Causality between taxes and expenditures: Evidence from Latin American countries," *Journal of Economics and Finance* **23**: 184–92.

Cuddington, J.T. (1997), "Analyzing the sustainability of fiscal deficits in developing countries," Policy Research Working Paper Series 1784. Washington, D.C.: The World Bank.

Edwards, S. (1995), *Crisis and reform in Latin America: From despair to hope.* Oxford: Oxford University Press.

Ehrhart, C. and M. Llorca, M. (2008), "The sustainability of fiscal policy: Evidence from a panel of six South-Mediterranean countries," *Applied Economics Letters* **15**: 797–803.

Engle, R.F. and C.W.J. Granger (1987), "Cointegration and error-correction: Representation, estimation, and testing," *Econometrica* **55**: 251–76.

Eslava, M. (2012), "The political economy of public spending and fiscal deficits: Lessons for Latin America," in J. Santiso and J. Dayton-Johnson, eds., *The Oxford Handbook of Latin American Political Economy.* Oxford: Oxford University Press.

Ewing, B. and J. Payne (1998), "Government tax revenue-expenditure nexus: Evidence from Latin America," *Journal of Economic Development* **23**: 57–69.

Filc, G. and C. Scartascini (2007), "Budget institutions," in E. Lora, ed., *The state of state reform in Latin America.* Stanford, California: Stanford University Press.

Friedman, M. (1978), "The limitations of tax limitation," *Policy Review* 5: 7–14.

Hadri, K. (2000), "Testing for stationarity in heterogeneous panel data," *Econometrics Journal* **3**: 148–61.

Hakkio, G. and M. Rush (1991), "Is the budget deficit too large?" *Economic Inquiry* **29**: 429–45.

Hallerberg, M. and C. Scartascini (2011), "Economic crisis and fiscal reforms in Latin America," Working Paper Series No. IDB-WP-235. Washington, D.C.: Inter-American Development Bank.

Hamilton, J. and M.A. Flavin (1986), "On the limitations of government borrowing: A framework for empirical testing," *American Economic Review* **76**: 808–19.

Haug, A.A. (1991), "Cointegration and government borrowing constraints: Evidence for the United States," *Journal of Business and Economic Statistics* **9**: 97–101.

Holmes, M.J., J. Otero and T. Panagiotidis (2010), "Are EU budget deficits sustainable?" *Empirical Economics* **38**:767–78.

Hoover, K.D. and S.M. Shefrin (1992), "Causation, spending and taxes: Sand in the sandbox or tax collector for the welfare state," *American Economic Review* **82**: 225–48.

Im, K., M. Pesaran and Y. Shin (2003), "Testing for unit roots in heterogeneous panels," *Journal of Econometrics* **115**: 53–74.

Johansen, S. (1988), "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control* **12**: 231–54.

Jones, J.D. and D. Joulfain (1991), "Federal government expenditures and revenue in the early years of the American Republic: Evidence from 1792-1860," *Journal of Macroeconomics* **13**: 133–55.

Joulfaian, D. and R. Mookerjee (1990), "The intertemporal relationship between state and local government revenues and expenditures: Evidence from OECD countries," *Public Finance* **45**: 109–17.

Kao, C. (1999), "Spurious regression and residual-based tests for cointegration in panel data," *Journal of Econometrics* **90**: 1–44.

Kao, C. and M.H. Chiang (2000), "On the estimation and inference of a cointegrated regression in panel data," *Advances in Econometrics* 15: 179–222.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and T. Shin (1992), "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic series have a unit root?" *Journal of Econometrics* **54**: 159–78.

Larsson, R., J. Lyhagen, and M. Löthgren (2001), "Likelihood-based cointegration tests in heterogenous panels," *Econometrics Journal* **4**: 109–42.

Lau, E. and A.Z. Baharumshah (2005), "Mean-reverting behavior of current account in Asian countries," *Economics Letters* **87**: 367–71.

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

MacDonald, R. (1992), "Some tests of the government's intertemporal budget constraint using U.S. data," *Applied Economics* **24**: 1287–92.

Maddala, G. and S. Wu (1999), "A comparative study of unit root tests and a new simple test," *Oxford Bulletin of Economics and Statistics* **61**: 631–52.

Manage, N. and M.L. Marlow (1986), "The causal relation between federal expenditures and receipts," *Southern Economic Journal* **52**:617–29.

Miller, S.M. and F.S. Russek, "Cointegration and error-correction models: temporal causality between government taxes and spending," *Southern Economic Journal* **57**: 33–51.

Munawar-Shah, S. and M. Abdul-Majid (2014), "Assessing fiscal sustainability for SAARC and IMT-GT Countries," *Romanian Journal of Fiscal Policy* **5**: 26–40.

Musgrave, R. (1966) "Principles of budget determination," in Cameron, H. and W. Henderson eds., *Public finance: Selected readings*. New York: Random House.

Nyamongo, M.E., M.M. Sichei and N.J. Schoeman (2007), "Government revenue and government expenditure nexus in South Africa," *SAJEMS* NS **10**: 256–68.

Ocampo, J. (2009), "Latin America and the global financial crisis," *Cambridge Journal of Economics* **33**: 703–24.

Owoye, O. (1995), "The causal relationship between taxes and expenditures in the G7 countries: Co-integration and error correction models," *Applied Economic Letters* 2: 19–22.

Papadopoulos, A. and M. Sidiropoulos (1999), "The sustainability of fiscal policies in the European Union," *International Advances in Economic Research* **5**: 289–307.

Peacock, A.T. and J. Wiseman (1979), "Approaches to the analysis of government expenditures growth," *Public Finance Quarterly* **7**: 3–23.

Pedroni, P. (1999), "Critical values for cointegrating tests in heterogeneous panels with multiple regressors," *Oxford Bulletin of Economics and Statistics* **61**: 653–70.

Pedroni, P. (2004), "Panel cointegration: Asymptotic and finite sample properties of pooled time series tests with an application to the purchasing power parity hypothesis," *Econometric Theory* **20**: 597–625.

Perron, P., "Dealing with structural breaks," in T. C. Mills and K. Patterson, eds., *Handbook of Econometrics*, Volume 1. London: Palgrave Macmillan.

Pesaran, H., Y. Shin and R. Smith (1999), "Pooled mean group estimation of dynamic heterogeneous panels," *Journal of the American Statistical Association* 94: 621–34.

Phillips, P. and P. Perron (1988), "Testing for a unit root in time series regression," *Biometrika* 75: 335–46.

Prohl, S. and F. Schneider (2006), "Sustainability of public debt and budget deficit: Panel cointegration analysis for the European Union member countries," Working Paper No 0610, Johannes Kepler University Linz.

Quantitative MicroSoftware (2014), EViews 8.1 User's Guide II. Irvine: New York

Quintos, C. (1995), "Sustainability of the deficit process with structural shifts," *Journal of Business & Economic Statistics* **13**: 409–17.

Ram, R. (1988), "A multicountry perspective on causality between government revenue and government expenditure," *Finance Publiques* **2**: 261–69.

Ross, K.L. and J.E. Payne (1998), "A re-examination of budgetary disequilibria," *Public Finance Review* **26**: 69–79.

Saikkonen, P. (1991), "Asymptotically efficient estimation of cointegration regressions," *Econometric Theory* **7**: 1–21.

Sanchez, O. (2003), "Globalization as a development strategy for Latin America?" *World Development* **31**: 1977–95.

Shinnick, E., ed., (2008), *Public finance, monetary policy, and market issues.* Berlin: Transaction Publishers.

Stock, J.H. and M.W. Watson (1993), "A simple estimator of cointegrating vectors in higher order integrated systems," *Econometrica* **61**: 783–820.

Trehan, B. and C.E. Walsh (1988), "Common trends, the government's budget constraint, and revenue smoothing," *Journal of Economic Dynamics and Control* **12**: 425–44.

Vanhorebeel, F. and P. van Rompuy (1995), "Solvency and sustainability of fiscal policies on the EU," *De Economist* **143**: 457–73.

Wagner, R.E. (1976), "Revenue structure, fiscal illusion, and budgetary choice," *Public Choice* **25**: 30–61.

Westerlund, J. (2007), "Testing for error correction in panel data," *Oxford Bulletin of Economics and Statistics* 69: 709–48.

Westerlund, J. (2008), "Panel cointegration tests of the fisher hypothesis," *Journal of Applied Econometrics* **23**: 193–233.

Westerlund, J. and S. Prohl (2010), "Panel cointegration tests of the sustainability hypothesis in rich OECD countries," *Applied Economics* **42**: 1355–64.

Westerlund, J. (2005), "New simple tests for panel cointegration," *Econometric Reviews* **24**: 297–316.

Westerlund, J. and D. Edgerton (2007), "A panel bootstrap cointegration test," *Economics Letters* **97**: 185–90.

Wilcox, D. (1989), "The sustainability of government deficits: implications of the present-value borrowing constraint," *Journal of Money, Credit and Banking* 21: 291–306.

World Bank (2015), *World Development Indicators.* Washington, D. C.: World Bank.