William H. Bowser

The long and short of returns to public investments in fifteen Ethiopian villages

February 2015

Replication Paper 4

Agriculture and Transportation

International Initiative for Impact Evaluation

About 3ie

3ie is an international grant-making NGO promoting evidence-informed development policies and programmes. We are the global leader in funding and producing high-quality evidence of what works, how, why and at what cost. We believe that better and policy-relevant evidence will make development more effective and improve people's lives.

3ie Replication Paper Series

The 3ie Replication Paper Series is designed to be a publication and dissemination outlet for internal replication studies of development impact evaluations. Internal replication studies are those that reanalyse the data from an original paper in order to validate the results. The series seeks to publish replication studies with findings that reinforce an original paper, as well as those that challenge the results of an original paper. To be eligible for submission, a replication study needs to be of a paper in 3ie's online Impact Evaluation Repository and needs to include a pure replication. 3ie invites formal replies from the original authors. These are published on the 3ie website together with the replication study.

The **3ie Replication Programme** also includes grant-making windows to fund replication studies of papers identified on the candidate studies list. Requests for proposals are issued one to two times a year. The candidate studies list includes published studies that are considered influential, innovative or counterintuitive. The list is periodically updated based on 3ie staff input and outside suggestions. The aim of the 3ie Replication Programme is to improve the quality of evidence from development impact evaluations for use in policymaking and programme design.

About this report

This paper was funded through 3ie's Replication Window with generous funding from an anonymous donor. All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or the 3ie Board of Commissioners. Any errors and omissions are the sole responsibility of the authors. Comments and queries should be directed to the corresponding author, William H Bowser, at William_Bowser@abtassoc.com.

Suggested citation: Bowser, WH 2015. *The long and short of returns to public investments in fifteen Ethiopian villages.* 3ie Replication Paper 4. Washington, DC: International Initiative for Impact Evaluation (3ie).

3ie Replication Paper Series executive editor: Annette N Brown
Managing editor: Benjamin DK Wood
Assistant managing editor and production manager: Jennifer Ludwig
Layout Assistant: Kara Ingraham
Copy editor: Pamela Tatz
Cover design: John F McGill
Proof reader: Yvette Charboneau

The long and short of returns to public investments in fifteen Ethiopian villages

William H Bowser

3ie Replication Paper 4
February 2015

Acknowledgments

I would like to thank the original authors of the study under replication for sharing the data and program code used to conduct this research. I would like to especially thank John Hoddinott for his enthusiasm regarding the completion of this replication. I thank the International Initiative for Impact Evaluation and their excellent team of Replication Managers for their support in executing this replication. I would also like to thank Alan de Brauw for providing excellent comments during the external review. All remaining errors are my own.

Abstract

Public investments play an important role in transforming poverty cycles for rural households. Since public-investment portfolios are often diversified to more effectively allocate scarce resources to meet a variety of needs, it is perhaps more critical for policymakers to know which investments yield the greatest returns towards welfare objectives. This study replicates research on a set of rural Ethiopian households conducted with the objective of quantifying the impact of public investments on the livelihoods of the rural poor. The pure replication supports the original authors' findings under the central assumption that household consumption levels, capital stock and access to technology change slowly over time. This replication also presents an alternative interpretation of the original qualitative results, with new quantitative evidence on the impact of investments in road infrastructure and the provision of agricultural extension service. Road development is shown to have a direct impact on short-term consumption and crop income growth, though there is weaker evidence of a long-run effect. However, extension services are shown to yield no impact in the short term or in the long run. The approach to the replication is supported by validation tests for the key underlying assumptions made in the original study and includes an application of an improved dynamic panel Generalized Method of Moments (GMM) estimator, which corrects for the bias in standard error estimates generated by the efficient two-step GMM estimator used to draw inferences in the original work. The replication findings suggest high economic returns to further investments in road infrastructure but indicate serious shortcomings in the effectiveness of extension services to support consumption and crop income growth for the rural households studied over the sample period. These results highlight public investments in Ethiopia that are likely to be highly effective when scaled nationally versus investments that appear too inefficient to successfully scale under the current capacity.

Contents

List of tables

Abbreviations and acronyms

GoE	Government of Ethiopia
PASDEP	Plan for Accelerated and Sustained Development to End Poverty
GMM-IV	Generalized Method of Moments Instrumental Variable
ERHS	Ethiopian Rural Household Survey
AEU	Adult-Equivalent Unit
PA	Peasant Association
OLS	Ordinary Least Squares
FE	Fixed Effects

1. Why replicate?

Dercon *et al*. (2009; DGHW, henceforth) provide crucial insight into the value of public investments on poverty and household well-being. Within many contexts, this study lays the foundation for important extensions into policy design. Notably, the mixed look at rural infrastructure developments and agricultural extension services corresponds with the multidimensional approach to public-resource allocation. Mogues (2011) argues that public-investment decisions, especially in the developing world, are made by necessity across multiple sectors. DGHW present a rare account of the impact of a diversified public-investment strategy that reaches policymakers on familiar terms and enriches the understanding for the development community.

Moreover, the focus on rural infrastructure projects in a localised setting provides a basis for forming expectations about the impact of developments with a broader reach, such as regional or national highway projects. Such broader development projects are natural extensions to localised investments (BenYishay and Tunstall 2011). Evidence of these successes on relatively small scales can support the advancement of initiatives for scaling up, which can simultaneously impact a larger populace.

Additionally, the provision of agricultural extension services generates a channel for productivity gains that may otherwise be foregone due to the difficulties involved in establishing a market for agricultural information diffusion. Agricultural extension services are characterised by their public-good nature, which makes efficient pricing a challenge, despite their obvious value (Maffioli *et al*. 2011). It is generally understood that growth in the agricultural sector of an economy will stimulate growth in other sectors (Dercon and Zeitlin 2009). Ultimately, the goal of public investments such as agricultural extension services is to accelerate growth in agricultural production and overall incomes, especially for agricultural smallholders. DGHW illustrate this point by uniquely characterising the impact of access to agricultural extension services in terms of consumption growth for smallholders. This is an important perspective to take, particularly in Ethiopia, where the Government of Ethiopia's (GoE) Plan for Accelerated and Sustained Development to End Poverty (PASDEP) points to the agricultural sector to spearhead its economic growth, with a special focus on enhancing the position of smallholders in the sector (Dercon and Zeitlin 2009).

The emphases on investments that impact the rural segment of the country are made transparent by the rapid growth of investments dedicated to agriculture and road construction. Between 1999 and 2008, the GoE increased allocations to rural investments by 97.5 per cent, with 23.8 per cent of its expenditures constituting investments in rural areas (Dorosh and Schmidt 2010). Hence, the findings by DGHW documenting the experience in Ethiopia are likely to influence future policy designs both in Ethiopia and in other developing countries. Therefore, validation of the findings can enhance confidence in the resulting policy trajectory.

2. Pure replication

DGHW adapt a reduced-form economic growth model into a model depicting the growth path of household consumption as a function of access to technology (extension services), capital stock (road infrastructure) and transitory shocks.[1] Levels of consumption are measured by the household's reported food and nonfood expenditures, excluding durable goods related to investment (for example, education and healthcare expenditures). Additionally, the authors estimate the impact of receiving extension services and having access to all-weather roads on household-level poverty. There are two potential sources of endogeneity in their empirical model that must be addressed. To do so, DGHW employ a Generalized Method of Moments Instrumental Variable (GMM-IV) estimator with corrections for household-level fixed effects due to unobserved time-invariant heterogeneity in their sample. DGHW apply an important underlying assumption in the specification of their empirical growth model. Specifically, they assume that log consumption, access to technology and capital stocks change very slowly, such that the initial period levels are approximately equal to the observed levels in all subsequent periods. This assumption enables the authors to measure the growth process as an average growth rate across the entire sample period, while simplifying the identification of the parameter estimates on the explanatory variables. The need for simplifying the parameter estimates arises due to the uneven spacing between the periods in the sample.

Using data from the Ethiopian Rural Household Survey (ERHS), the main findings of the original study reveal that public investments in the form of extension services and improved roads can reduce poverty and increase economic growth for rural households. Specifically, headcount poverty in their sample is reduced by 9.8 percentage points from receiving at least one visit from an extension officer, while access to all-weather roads reduces poverty by 6.9 percentage points. Receiving at least one visit from an extension officer is also shown to increase consumption growth by 7.1 per cent, while access to all-weather roads increases consumption growth by 16.3 per cent. These results are encouraging and provide support for policies emphasising the role of public investments in economic growth and poverty alleviation.

DGHW provide an extensive array of robustness checks for their estimator: testing the sensitivity to weak instruments with Limited Maximum Likelihood estimation results, testing the sensitivity to outliers; treating access to technology as endogenous (adding the lag number of extension agents to the instrument set), additional controls for household characteristics and functional forms. They also disaggregated the estimated impacts through stratification on household characteristics. Generally, these robustness checks seem to imply a conservative estimate of the impact from public investments, based on their main results.

The pure replication of this study begins with the construction of key variables in the main regressions. The key outcomes measured include the headcount-based measure of the poverty rate for a household in a given survey round as well as a measure of real consumption growth per adult-equivalent unit (AEU), based on the consumption of food items and other nondurable goods. Both the poverty headcount and the real

[1] Note that the reference to capital stocks here pertains to the public accumulation of physical capital in the form of road infrastructure. Household access to road infrastructure is facilitated by public investments, rather than private capital accumulation.

consumption per adult-equivalent variables are provided in the data files supplied by DGHW. The main explanatory variables of interest are a dummy variable indicating whether a household has received at least one visit by an extension agent and another dummy variable indicating whether households within the Peasant Association (PA) have access to all-weather roads. Receiving a visit by an extension agent is used as a proxy for household access to technology. DGHW assume that extension services are the key channel by which households gain access to technologies for agricultural production. Furthermore, DGHW consider access to road infrastructure as the household's exogenous levels of capital stocks (Dercon *et al*. 2009). The definition of accessible roads in the ERHS questionnaire is important to the interpretation of the impact of road infrastructure. The question regarding the quality of roads is posed in the community- or PA-level questionnaire to a community representative in the section entitled 'Location and Access'. Questions 7 and 8 ask, respectively, 'How good is access via this road to and from the village in the rainy season?' and 'How good is access via this road to and from the village outside the rainy season?' Respondents select from the following set of options: 'WELL ACCESSIBLE TO ANY VEHICLES'; 'REASONABLE ACCESS TO ANY VEHICLES'; 'GOOD ACCESS TO TRUCKS AND BUSES'; 'REASONABLE ACCESS TO TRUCKS AND BUSES'; 'ACCESS TO CARTS/ANIMALS'; and 'ONLY WALKING'. DGHW define an all-weather road as one that is either 'well accessible to any vehicles' or allows 'reasonable access to any vehicle' during the rainy season. As the judgment of 'reasonable access' marks the threshold that qualifies roads surrounding the PA, the definition of all-weather roads ends up being highly subjective, with the potential to overestimate the number of households with access to quality road infrastructure. This may justify alternative definitions of all-weather roads in extensions of the current replication.

The additional covariates include the lagged level of consumption per adult in each household, the log change in rainfall since the previous round, the change in the log output price index for each PA since the previous round, indicator variables for negative input price shocks, death and sickness shocks since the last round, and a dummy variable indicating whether the interview was conducted during the post-harvest season. Because of the uneven time spacing between rounds, the scaling of the variables by the length (in years) between survey rounds is instrumental in constructing the variables deployed for the estimation.

The cooperation of the original authors, who graciously provided the codes and data for this replication, facilitated the construction of the variables. The construction of most variables is transparent. The key instrumental variables are employed to control for the endogeneity of the lagged dependent variable (most important in the context of the consumption growth model). As noted in the original publication, 'instruments for lagged endogenous variables are lagged log livestock units per adult equivalent, lagged log number of adult equivalents, and lagged log cultivable land per adult equivalent'. These values are straightforward to calculate from the data files, and DGHW create these instruments cleanly, except in the case of the log number of AEUs in the household. Apparently, DGHW omit the natural logarithmic transformation of the AEU variable in their code. Because this is a key instrumental variable, I update this transformation in the replication code.

For the purposes of the pure replication, I execute the GMM-IV estimation on the unevenly spaced panel given the above construction of the variables, still controlling for household-level fixed effects. The results of this stage of the estimation correspond to DGHW's table 3 and are summarised here in table 1. The parameter estimates on the variables of interest are qualitatively consistent with the results reported by DGHW, although with small deviations. The full model estimating the impact of public investments on the poverty rate, with controls for changes in rainfall, output prices and household shocks, is reported in column [2]. Receiving at least one visit from the extension agent in the previous round is shown to reduce the poverty rate by 9.84 percentage points, while having access to all-weather roads in the previous round is shown to reduce poverty by 6.86 percentage points. The full model specification estimating the impact on real consumption growth is reported in column [4]. Households receiving at least one visit from an extension agent in the previous round report an average growth rate 7.28 percentage points above those households without access to extension services. Having access to all-weather roads enables households to realise an average growth rate 16.2 percentage points above those households lacking access to more reliable road infrastructure. These values are in alignment with the original findings of DGHW, with small deviations likely due to the natural logarithmic transformation of the AEU instrumental variable. Similarly, the diagnostic tests on the instruments support the relevance of the chosen instrument set by DGHW.

These results show that under the key assumptions of slow changes in levels of capital stock and access to technology, the original authors' conclusions are indeed robust to replication, despite a slight programming oversight. The next stage of the replication will endeavour to further explore the validity of the assumption made, which enables the construction of *p*-period averages across the unevenly spaced rounds of the survey.

Table 1: Pure replication results: determinants of poverty and consumption growth

	[1] Poor	[2] Poor	[3] Consumption growth	[4] Consumption growth
Lagged log consumption	− 0.138***	− 0.128***	− 0.394***	− 0.406***
	(0.0359)	(0.0358)	(0.0384)	(0.0383)
Received at least one extension visit (previous round)	− 0.0962**	− 0.0984***	0.0615*	0.0728**
	(0.0381)	(0.0380)	(0.0364)	(0.0361)
Access to all-weather roads (previous round)	− 0.0679**	− 0.0686**	0.165***	0.162***
	(0.0315)	(0.0316)	(0.0303)	(0.0303)
Log change in annual rainfall since last round		− 0.0258		0.114***
		(0.0323)		(0.0379)
Change in log output price index since last round		0.0922***		− 0.172***
		(0.0185)		(0.0225)
Negative input price shock since last round; yes = 1, no = 0		− 0.0533		− 0.117
		(0.113)		(0.116)
Death shock since last round; yes = 1, no = 0		0.0778		− 0.162**
		(0.0710)		(0.0770)
Illness shock since last round; yes=1, no=0		0.0831		− 0.107
		(0.0831)		(0.0890)
Post-harvest season at time of interview; yes=1, no=0	0.0881***	0.0907***	0.104***	0.116***
	(0.0163)	(0.0165)	(0.0157)	(0.0152)
Diagnostic statistics				
Observations	4,781	4,771	4,781	4,771
Kleibergen-Paap Wald rank F statistic	110.452**	109.81**	110.452**	109.807**
Hansen J-test	0.575	0.766	3.609	4.205

Notes: Robust standard errors are in parentheses. Columns [1] and [2] estimate a GMM-IV model on the poverty headcount conditional on the listed covariates. Columns [3] and [4] estimate the same GMM-IV model on the change in household consumption growth. The endogenous variable is lagged log consumption. Instrumental variables include log adult-equivalent units per household, log land holdings and log livestock units held per household – all with a single period lag.
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

3. Measurement and estimation analysis (MEA)

There are two main parts of the MEA section of my replication study. First, I examine the slow growth assumptions in regards to the technology, capital stock and consumption across the different data sets. After testing the validity of these assumptions, I explore alternative estimation strategies that only use the evenly spaced datasets. I then use these alternative models to estimate the short-run versus long-run effects of the intervention and test for additional endogenous variables.

3.1 Testing the slow growth assumptions

As previously mentioned, the empirical model employed by DGHW is dependent on the validity of the assumption that access to technology, capital stock accumulation and consumption levels change very slowly, such that the initial period level is approximately equal to the level in period $t - 1$. Revisiting the empirical growth model outlined by equation (4) in DGHW, we see that the p-period growth model should generally take the following form:

$$\frac{(y_t - y_{t-p})}{p} = \delta + \alpha \frac{(y_{t-1} + \ldots + y_{t-p})}{p} + \beta \frac{(k_{t-1} + \ldots + k_{t-p})}{p}$$

$$+\gamma \frac{(R_{t-1} - R_{t-p})}{p} + \lambda X + \varepsilon_t .$$

However, the authors maintain the assumption that $\ln y_{t\text{-}1} \approx \ln y_{t\text{-}2} \approx \ldots \approx \ln y_{t\text{-}p}$ and, similarly, $\ln k_{t\text{-}1} \approx \ln k_{t\text{-}2} \approx \ldots \approx \ln k_{t\text{-}p}$, which enables them to recharacterise equation (4) as

$$\frac{(y_t - y_{t-p})}{p} = \delta + \alpha \quad y_{t-p} + \beta \quad k_{t-p} + \gamma \frac{(R_{t-1} - R_{t-p})}{p} + \lambda X + \varepsilon_t .$$

This assumption can be tested formally by forming the null hypothesis that the mean difference in the levels of consumption, access to technology and capital stocks in period $t - 1$ and in period $t - p$ are equal to zero. In this application, I employ a multivariate test of mean equality across rounds for each variable. This multivariate test is effective in testing the mean equality across more than two groups (Mardia *et al*. 1976). Under the assumption that each variable was drawn from the same or a single sample, the Hotelling T-squared test is applicable, which reduces to a standard *t*-test in the case of a single variable. Because of the panel structure, the sample may change across rounds. Hence, the multiple sample test is also conducted with a likelihood-ratio test of mean equality, which allows for covariance heterogeneity across rounds. The results of the multivariate tests of these assumptions are reported below in table 2. In all cases, the null hypothesis is rejected, rendering the assumption invalid. This is not surprising, as the summary statistics reported by DGHW's table 2 indicate that the mean real consumption per adult equivalent grew by 36 per cent over the sample period. Over the same period, the number of households with access to all-weather roads increased by more than 30 percentage points, while access to extension services increased by 10 percentage points.

Table 2: Test of equality of means

H_0	One-sample test (Hotelling T-squared statistic)	Multiple-sample test (Likelihood-ratio)
Mean (Real consumption [t − 1] = ... = Real consumption [t − p])	311.38***	224.96***
Mean (Access to all-weather road [t − 1] = ... = Access to all-weather road [t − p])	143.93***	46.85***
Mean (Access to extension [t − 1] = ... = Access to extension [t − p])	93.61***	57.97***

Note: *** indicates significance at the 1 per cent level

As we are no longer able to assume that consumption, access to technology and capital stocks evolve slowly over the sample period, the appropriate specification would require taking p-period averages of the data and applying the DGHW estimator. Results of this procedure are reported in table 3. The estimation results in column [1] still suggest that access to all-weather roads reduces the poverty rate and that it does so by nearly three times the rate of the estimates reported by DGHW. However, the effect of access to technologies, for which the authors used extension visits as a proxy, on the poverty rate is no longer statistically significant, despite retaining the same sign on the coefficient. The results of the consumption growth estimation are also disappointing under this specification. Neither of the variables of interest is statistically significant.

Table 3: Measurement and estimation analysis replication results (p-period averages)

Dependent Variables	[1] Poor	[2] Consumption growth
Log consumption (period $t-1$ to $t-p$ average)	0.0457***	− 0.0156
	(0.0101)	(0.0120)
Received at least one extension visit (period $t-1$ to $t-p$ average)	− 0.0579	0.0723
	(0.0666)	(0.0891)
Access to all-weather roads (period $t-1$ to $t-p$ average)	− 0.185***	− 0.0110
	(0.0472)	(0.0548)
Log change in annual rainfall since last round	0.0410	0.261***
	(0.0284)	(0.0451)
Change in log output price index since last round	0.0998***	− 0.182***
	(0.0175)	(0.0292)
Negative input price shock since last round; yes = 1, no = 0	− 0.0355	− 0.0987
	(0.106)	(0.159)
Death shock since last round; yes = 1, no = 0	0.0844	− 0.204*
	(0.0652)	(0.105)
Illness shock since last round; yes = 1, no = 0	0.0279	− 0.111
	(0.0779)	(0.115)
Post-harvest season at time of interview; yes = 1, no = 0	− 0.115***	0.140***
	(0.0171)	(0.0234)
Diagnostic statistics		
Observations	4,959	4,959
Kleibergen-Paap Wald rank F statistic	1522.789**	1522.789**
Hansen J-test	1.896	7.015

Notes: Robust standard errors are in parentheses. Column [1] estimates a GMM-IV model on the poverty headcount conditional on the listed covariates. Column [2] estimates the same GMM-IV model on the change in household consumption growth. Lagged log consumption is the endogenous variable. Instrumental variables include log adult-equivalent units per household, log land holdings and log livestock units held per household, all with a single-period lag.
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

3.2 Dynamic GMM

Andreou *et al*. (2010) demonstrate that using equal weights to aggregate data collected with different frequencies may lead to inefficient, biased and inconsistent parameter estimates. Hence, using the p-period averages may produce inconsistent results. The p-period averages used in the above specification of the consumption growth model are based on equal weights across each sampling period. This explains the surprising results found in table 3. Rather than pursuing the nonlinear least squares estimator developed to deal with the Mixed Data Sampling problem alluded to by Andreou *et al*. (2010), I consider estimators best suited to model data sampled across evenly spaced intervals.

I explore possibly more efficient modeling techniques by using only the evenly spaced data. This approach requires dropping some of the data sets, but it allows for a consistent approach to the estimation strategy. Evenly spaced data enables the use of dynamic panel GMM estimators, that is, first-difference and/or system GMM (Arellano and Bond 1991; Blundell and Bond 1998, 2000; Bond *et al.* 2001). Dynamic panel GMM estimators are best suited in situations where the panel is wide ($N \rightarrow \infty$) with a short and finite time dimension ($T \geq 3$). In typical applications, the model contains a single dependent lagged variable, with some set of control or policy variables, and includes time-invariant fixed effects. Consider the following model:

$$y_{it} = \psi y_{i,t-1} + x'_{it}\beta + \varepsilon_{it}$$

$$\varepsilon_{it} = \mu_i + \nu_{it} \tag{1}$$

$$E[\mu_i] = E[\nu_{it}] = E[\mu_i \nu_{it}] = 0$$

In the above model, i denotes the observational unit of the N-dimensional panel and t denotes the time period of observation, where T is the maximum number of periods in the sample. The outcome variable is expressed as y, and let $\boldsymbol{x}$ be the vector of control and/or policy variables that may contain lagged variables as well. The error term is composed of the fixed effects μ_i and the idiosyncratic disturbance term v_{it}. As in DGHW, equation (1) can be written as a growth equation of the following form:

$$\Delta y_{it} = \alpha y_{i,t-1} + x'_{it}\beta + \varepsilon_{it} \tag{2}$$

The difference GMM eliminates the fixed effects nuisance parameter by taking first differences of the data and estimating a linear GMM model of the following form:

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta \boldsymbol{x}'_{it}\beta + \Delta \nu_{it} \tag{3}$$

This transformation eliminates the fixed effects but does not correct for the endogeneity of the lagged dependent variable. Following the work of Holtz-Eakin *et al*. (1988), the instrument set for the lagged endogenous variable (in differences) can be constructed with lags of the endogenous variable in levels, because lagged levels are independent of the differenced contemporary error term (that is, $E\ [y_{i,t-1}\Delta \mathrm{v}_{\mathrm{it}}] = 0$ for each $t \geq 3$, $l \geq 2$) as long as there is no serial correlation in the idiosyncratic errors. Similarly, in the case of predetermined covariates $\omega \subseteq \boldsymbol{x}$, lags of ω can also be used as instruments, because $E\ [\omega_{i,t-1}\Delta \mathrm{v}_{\mathrm{it}}] = 0$ for each $t \geq 2$, $l \geq 1$. First differences of additional strictly exogenous instruments, including strictly exogenous covariates, may also be included in the instrument set as standard instrumental variables.

The system GMM utilises a stacked set of the data, based on an equation for the levels and an equation for the first differences of the data. The instrument set is comprised of a set of lagged levels for the equation of differences and a set of lagged differences as instruments for the levels equation. So in addition to the moment conditions drawn from the difference GMM, the system GMM includes $E\ [y_{i,t-1}, \varepsilon_{\mathrm{it}}] = 0$. This moment condition stipulates that the instrument must be orthogonal to the fixed effects, as the untransformed error term in the levels equation contains the nuisance parameter μ_i. It can be shown that this condition will hold if the autoregressive process is convergent, or if $|\alpha| < 1$ (see Blundell and Bond [1998] for details).

The system GMM is more efficient than the difference GMM for a *y* series that is persistent or otherwise close to a random walk. However, in both the difference and the system GMM the instrument set is quadratic in *T* and can explode quite rapidly as the length of the panel increases. As the number of instruments approaches *N*, this can introduce bias in the fit of instrumented variables (especially one-step GMM), downwards bias in standard errors (especially two-step GMM) and weak tests of instrument validity (Roodman 2009b). However, it should be noted that a high instrument count does not make coefficient estimates from the two-step GMM estimator inconsistent. Though the efficiency of the two-step GMM is compromised as the instrument count increases, the correction developed by Windmeijer (2005) has been shown to improve inferences on standard errors drawn from two-step GMM.

Instrument proliferation can weaken the Hansen (1982) J-test for instrument validity as well as the difference-in-Hansen tests for the validity of instrument subsets. The bias generated by rapid growth of the instrument count as *T* grows has led many researchers to believe that the dynamic GMM estimation remains safe as long as the number of instruments is kept below *N* (Roodman 2009b). However, simulation results have shown that the integrity of the J-test is not necessarily maintained when the instrument count is kept below *N* and that implausibly high *p* values can be generated, leading to under-rejection of the null hypothesis that the instruments (and overall model specification) are valid (Andersen and Sørensen 1996; Bowsher 2002). This is particularly problematic in cases with small samples (short *N*) and more periods in the sample (long *T*).

One way of dealing with the issue of instrument proliferation is to reduce the number of lags used in the instrument set. However, this approach reduces the information available to fit the instrumented variable(s). A second approach is to combine or 'collapse' the instruments into smaller sets and impose the moment condition $E[y_{i,t-1}\Delta v_{it}] = 0$ for each $l \geq 2$. Collapsing the instrument set in this way reduces the burden of the estimator by requiring only the minimisation of the GMM criterion function for each *l*, rather than over both *l* and *t* (Roodman 2009b). Alternatively, both methods can be used jointly to check for improvements in the J-test. However, before proceeding further to illustrate the efficacy of dynamic panel GMM in this context, it is important to reconcile a few other specification issues that could potentially be problematic.

3.3 Short-run vs. long-run effects

The first issue at hand pertains to the entry of the capital stock and access to technology proxies in DGHW.[2] DGHW characterise these terms as single-period lagged variables in both the poverty and consumption growth models. This necessitates a different interpretation of the coefficient estimates than that which is alluded to in DGHW. The growth model employed by DGHW is equivalent to

$$y_t = \psi y_{t-1} + \beta x_t + \phi d_{t-1} + \varepsilon_t \tag{4}$$

[2] A minor observation worth noting is that the proxies representing k_{t-1} do not actually enter the estimation equation as natural logarithmic transformations, as depicted in equation (4) of Dercon *et al*. (2009). Such a transformation would render identification implausible, as both of these proxies enter as binary indicator variables.

where $\psi = \alpha + 1$ and $d_{t-1} = 1$ for units observed with treatment status in period $t - 1$ and $d_{t-1} = 0$ otherwise.[3] Observing treatment status in period $t - 1$ and measuring impact in period t should be interpreted as the long-term treatment effect. This specification implies that units move from nontreatment status in period $t - 2$ to treatment status in period $t - 1$, with outcomes measured in period t. In similar applications, long-term treatment effects have been found to reflect attenuated growth trajectories due to diminishing returns to investment (Khandker and Koolwal 2011). Whether or not diminishing returns to investment in public capital stocks exist is an important policy question of interest. However, ignoring the short-term effects in such a specification may result in misleading conclusions due to an omitted-variable bias.

While one could argue that the gains from public investments are realised with a lag, it seems plausible that both regressors of interest could have realised effects on consumption growth due to contemporary impacts on current income and expenditures. For example, access to markets via improved road infrastructure can result in higher incomes as well as greater consumption expenditure on goods due to increased mobility. Such gains can be realised during the same period that the public investment is observed. It is highly doubtful that it would take multiple years for a household member to benefit from improved road infrastructure or access to improved technologies.

Furthermore, policymakers may also be interested in learning about the effect of receiving the treatment in the near term (that is, observing treatment status in period t and measuring the outcome in the same period). Though attenuating returns are not necessarily the case at hand, it is of interest to specify a model that is conducive to identifying both short-term and long-term effects of having access to technology and all-weather roads.

3.4 Additional endogenous variables

The second issue at hand is related to potential sources of endogeneity other than the lagged dependent variable. Specifically, the treatment variables may be correlated with the error term in addition to the lagged dependent variable. DGHW indicate in their section on robustness checks that the instrument employed in the endogeneity correction for access to technology may be weak and lack sufficient coverage in the sample. This may warrant the use of alternative instrument(s) and/or methods to correct selection bias affecting the impact estimates.[4] Additionally, the authors mention but never directly address the issue of nonrandom road improvements. Because road development occurs at the village level in this study, it is difficult to disentangle the causal effect of capital stocks given the small number of clusters (villages) available to measure village-level effects. In Dercon *et al*. (2009), all-weather roads were made available to an increasing number of households and villages over the sample period. In this dynamic setting, lagged outcomes (and lagged realisations of the treatment variables) may be suitable instruments for potentially endogenous contemporary

[3] Henceforth, only the consumption growth model will be considered.

[4] DGHW note that the potential endogeneity of the regressors of interest may arise due to correlation with the time-invariant household characteristics. The dynamic GMM estimator also corrects for correlation with fixed effects, while enabling the use of alternative instruments drawn from lags of the regressors.

programme placement with the use of dynamic panel GMM estimators (Khandker and Koolwal 2011).[5]

As a first step toward investigating the potential endogeneity of the treatment variables, I test for endogeneity of access to extension and to all-weather roads under the GMM-IV specification used in DGHW on the evenly spaced panel. Specifically, this test takes the difference between the Sargan-Hansen statistic for the model treating the respective public-investment variable as endogenous and that for the model in which each suspect variable is treated as exogenous. The resulting test statistic has a chi-squared distribution and is also robust to violations of homoskedasticity (Baum *et al*. 2010). The results of the endogeneity tests are reported in table 4. For both access to extension and all-weather roads, the null hypothesis of exogeneity cannot be rejected. The exogeneity of these two variables is upheld for both the contemporaneous and lagged realisations. Hence, for now it is reasonably safe to consider lagged realisations of the treatment variables in the instrument set and treat contemporaneous realisations as exogenous.

Table 4: Endogeneity test of public investments (1994–2004 evenly spaced data), GMM FE

H_0	χ^2 *p* value
Access to extension in period $t-1$ can be treated exogenously	0.5939
Access to all-weather roads in period $t-1$ can be treated exogenously	0.2993
Access to extension in period t can be treated exogenously	0.2445
Access to all-weather roads in period t can be treated exogenously	0.6496

Note: The endogeneity test implemented is defined as the difference of two Sargan-Hansen statistics: one for the equation with the smaller set of instruments, where the suspect regressor(s) are treated as endogenous, and one for the equation with the larger set of instruments, where the suspect regressors are treated as exogenous.

4. Distinguishing between empirical models

Given these insights into the way both the data and the consumption growth model can be specified, we can also begin to examine more directly the efficacy of the first-differenced GMM estimator over the GMM-IV with fixed effects employed in DGHW. Bond (2002) explains that the coefficient estimate on the lagged dependent variable from pooled Ordinary Least Squares (OLS) and the Fixed Effects (FE) estimators can be used as respective upper and lower bounds on the range of reliable coefficient estimates in a panel setting. Credible estimates drawn from various theoretically superior models should report estimates of the lagged dependent variable within this range.

Table 5 reports the estimates of the consumption growth model, which measures the short- and long-term effects of having access to extension services and all-weather roads with OLS, standard FE and DGHW's GMM-IV estimator with FE. OLS and FE estimates are drawn from estimation of equation (1). Given the bounds suggested by columns [1] and [2], we should expect estimates on the lagged dependent variable to lie

[5] Lagged regressors are also commonly employed as instruments in the literature on foreign aid and growth (see, for example, Dalgaard *et al*. [2004] and Minoiu and Reddy [2010]) within dynamic GMM applications.

somewhere within the range of [−0.413, 0.213]. As DGHW estimate and interpret their model in the form of equation 2, we must add one to the estimated coefficient to recover an estimate comparable to columns [1] and [2]. Taking this into account, we see that the coefficient estimate for the autoregressive term drawn from DGHW's estimator reported in column [3] is within credible range (−0.199). However, the Hansen J statistic shows that under the given specification, the model is not overidentified. This implies that some or all of the instruments used to correct the endogeneity of the lagged dependent variable may be invalid.[6]

Further tests of the exogeneity of the instruments included in the DGHW model reveal that the lagged log AEU is not a valid instrument. The C statistic from the corresponding difference in Sargan-Hansen statistic is reported in column [3]. Column [4] reports the estimation results of the DGHW model, where only lags of log livestock units held per adult equivalent and log land size per adult equivalent are included in the instrument set. In this case, the Hansen J statistic indicates that the model is overidentified and the instruments are valid. However, the coefficient estimate for the autoregressive term indicates that the estimator generates estimates with a downwards bias. Adding one to the coefficient yields an autoregressive estimate of −0.417, which is just outside the lower bound on the credible range. Hence, one must take caution in drawing inference on the causal effect of access to extension and access to all-weather roads under this specification, as there appears to be a bias that is slightly worse than the bias present in the FE results.

The distinction between the GMM model employed by DGHW and the difference GMM can be understood as a comparison of different instrument sets or moment conditions. Both estimators eliminate the endogeneity attributable to the fixed effects, but the dynamic panel bias remains after making the correction for fixed effects. To correct for the endogeneity of the dynamic variable, the DGHW estimator relies on external or excludable instruments available outside the empirical dataset, whereas the difference GMM relies on transformations of the lagged dependent variable (and any other additional strictly exogenous variables). This approach to constructing instruments for the difference GMM is most convenient when no other excludable instruments are available.

Another important distinction between the DGHW GMM-IV model and the difference GMM is the treatment of the second moments. Though the efficient two-step GMM generates consistent coefficient estimates (first moments), the asymptotic standard errors are known to be small (Arellano and Bond 1991; Windmeijer 2005). This could lead to the over-rejection of the null hypothesis. The difference GMM employed in this replication corrects for the inconsistent standard errors derived from the efficient two-step GMM via the Windmeijer (2005) approach, whereas the DGHW estimator relies on uncorrected standard errors.

[6] Recall that the primary results in DGHW use lags of log livestock units held per adult equivalent, log land size per adult equivalent, and log adult equivalent as excluded instruments in the instrument set to control for the endogeneity of lagged log consumption in the growth equation.

Table 5: Determinants of consumption growth using OLS, Fixed Effects and GMM-IV Fixed Effects on evenly spaced data (1994–2004)

Dependent variable: *ln*(real consumption$_{it}$)	[1] OLS	[2] Fixed effects	[3] DGHW GMM-IV, FE	[4] DGHW GMM-IV, FE
log consumption [t − 5]	0.213***	− 0.413***	− 1.199***	− 1.417***
	(0.0185)	(0.0238)	(0.0669)	(0.105)
received at least one extension visit [t]	0.120***	0.181***	0.189***	0.170***
	(0.0423)	(0.0596)	(0.0603)	(0.0588)
received at least one extension visit [t − 5]	− 0.0239	0.0843	0.0581	0.0670
	(0.0505)	(0.0737)	(0.0786)	(0.0761)
access to all-weather roads [t]	0.479***	0.573***	0.543***	0.547***
	(0.0371)	(0.0530)	(0.0590)	(0.0650)
access to all-weather roads [t − 5]	0.0536	0.0378	0.0455	0.0143
	(0.0352)	(0.0550)	(0.0573)	(0.0584)
Δ annual rainfall since last round	0.465***	0.459***	0.484***	0.508***
	(0.0653)	(0.0759)	(0.0751)	(0.0747)
Δ output price index since last round	− 0.0357	0.137***	0.122***	0.110***
	(0.0299)	(0.0313)	(0.0334)	(0.0353)
negative input price shock since last round; yes = 1, no = 0	− 0.0693	0.115	0.0943	0.0789
	(0.0707)	(0.0815)	(0.0906)	(0.0870)
death shock since last round; yes = 1, no = 0	0.0427	− 0.00942	0.0246	− 0.00139
	(0.0565)	(0.0627)	(0.0683)	(0.0661)
illness shock since last round; yes = 1, no = 0	− 0.0517	− 0.0182	0.000516	− 0.0297
	(0.0554)	(0.0652)	(0.0685)	(0.0666)
post-harvest season at time of interview; yes = 1, no = 0	0.232***	− 0.137***	− 0.0592	− 0.0688
	(0.0321)	(0.0329)	(0.0405)	(0.0695)
Constant	3.096***	5.777***		
	(0.0804)	(0.113)		
Diagnostics				
Observations	2,494	2,494	2,324	2,324
Kleibergen-Paap Wald rank F statistic			61.759***	27.816***
Hansen J-test statistic			5.447*	0.656
C statistic (exogeneity test of lagged log adult-equivalent units)			5.684**	

Note: Robust standard errors are in parentheses. All empirical models estimate log real consumption growth conditional on the covariates. Column [1] estimate a Pooled Ordinary Least Squares model. A fixed effects model is specified in column [2]. Columns [3] and [4] estimate the same GMM-IV model found in DGHW. Instrumental variables include log adult-equivalent units per household, log land holdings and log livestock units held per household, all with a single-period lag. Column [4] uses only lags of log livestock units held per adult equivalent and log land size per adult equivalent as instrumental variables.
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

4.1 Checking robustness with dynamic GMM

Given the current instrument set employed in the DGHW model reported in table 5, access to extension services seem to yield a 17 percentage point short-run return to growth, with no long-run impact. Furthermore, this model suggests that access to all-weather roads increases consumption growth by nearly 55 percentage points in the short run, with no long-run impact on consumption growth after five years. However, this instrument set produces a dubious coefficient estimate of the autoregressive term (that is, the consumption growth rate), which brings the specification, including the current instrument set, into question. Again, doubt is cast over this model's reliability due to the apparent downwards bias in the autoregressive coefficient estimate reported in column [4].

Hence, we can check the robustness of the reported impact of public investments, and the overall model specification, by considering the alternate moment conditions generated by the difference GMM. The first-differenced GMM estimation results are generated from a model based on equation (1) and are presented in table 6.[7] Column [1] reports estimation results based on a model that employs the same set of controls as those used in DGHW, plus a set of year dummy variables. The instrument set includes the standard excluded instruments of first-differenced lagged log livestock units per AEU and lagged log land holdings per AEU, each under the assumption of strict exogeneity, as well as all included variables self-instrumenting themselves with first differences. Additionally, a set of 'GMM style' instruments are employed, including the second-order lag of the dependent variable and all available lags of the treatment variables (that is, lagged access to extension and all-weather roads).[8]

Columns [1] through [3] report estimates based on the collapsed set of GMM-style instruments. Columns [2] and [3] report estimates from a more parsimonious model, which excludes the covariates with no explanatory power initially considered by DGHW, and the model in column [1]. As previously mentioned, all first-differenced GMM estimates use the robust two-stage estimation procedure to calculate the variance and make corrections to the standard errors according to the approach developed by Windmeijer (2005).

Each of the difference GMM models report coefficient estimates on the lagged dependent variable that are now within the credible range (that is, the bounds of the autoregressive parameter estimate generated by OLS and FE). The specifications in columns [1] and [2] each use 14 instruments, while the model summarised in column [3] uses 13 instruments. The instrument count across each specification is low relative to the sample size.[9]

[7] First-differenced GMM estimation is executed in Stata using the *xtabond2* syntax developed by Roodman (2009a).
[8] GMM-style instruments are the set of lagged dependent variables built in a matrix structure.
[9] Given the number of regressors, the model needs a minimum of 13 instruments in the full specification and nine instruments in the parsimonious specifications to be overidentified. Hence, instrument proliferation is not a concern in either specification.

Table 6: Determinants of consumption growth using first-differenced GMM on evenly spaced data (1994−2004)

Dependent variable: *ln*(real consumption$_{it}$) or *ln*(real crop income$_{it}$)	[1]	[2]	[3]	[4]
	Diff. GMM	Diff. GMM	Diff. GMM	Diff. GMM
log consumption (or crop income) [t − 5]	− 0.223***	− 0.113***	− 0.409***	− 0.328***
	(0.0653)	(0.0376)	(0.106)	(0.0686)
received at least one extension visit [t]	0.0882	0.160	− 0.124	0.0578
	(0.166)	(0.164)	(0.171)	(0.202)
received at least one extension visit [t − 5]	− 0.0561	− 0.0376	− 0.0945	− 0.163
	(0.118)	(0.128)	(0.106)	(0.143)
access to all-weather roads [t]	0.723***	0.756***	0.455***	0.329*
	(0.140)	(0.116)	(0.154)	(0.174)
access to all-weather roads [t − 5]	0.114	0.135*	− 0.00100	− 0.0608
	(0.0769)	(0.0692)	(0.0782)	(0.0947)
Δ annual rainfall since last round	0.589***	0.620***	0.450***	0.176
	(0.0989)	(0.0957)	(0.108)	(0.157)
Δ output price index since last round	0.159***	0.153***	0.0775	0.290***
	(0.0498)	(0.0408)	(0.0478)	(0.0671)
negative input price shock since last round; yes = 1, no = 0	− 0.0511			
	(0.0684)			
death shock since last round; yes = 1, no = 0	0.0508			
	(0.0918)			
illness shock since last round; yes = 1, no = 0	0.0342			
	(0.0689)			
post-harvest season at time of interview; yes = 1, no = 0	− 0.00915			− 0.529***
	(0.0690)			(0.0790)
Diagnostics				
Observations	1,162	1,162	1,162	975
Number of Instruments	14	14	13	13
Hansen J-test statistic	5.06*	8.77	1.70	7.34

Note: Robust standard errors are in parentheses. Columns [1] and [2] estimate real consumption growth with a first differenced GMM model, while columns [3] and [4] estimate real crop income growth via the same empirical model.
*** p<0.01, ** p<0.05, * p<0.1

The parsimonious models include the transitory input, death and illness shocks as well as the seasonal timing of the survey as excludable and strictly exogenous instruments, in addition to the DGHW instruments and the lags of the endogenous dependent variable used as instruments. Perhaps the most noteworthy benefit to placing these variables among the excluded instruments is marked by the improvement in the model validity as reported by the Hansen J-test statistic. The model in column [1] rejects the validity of the overidentifying restrictions at the 5 per cent significance level, while the validity of the model is not rejected at any conventional level of significance under the more parsimonious specifications reported in columns [2] and [3].

The validity of the second-order lag of the dependent variable as an instrument is compromised in the presence of serial correlation in the levels of the idiosyncratic disturbances. Typically this can be verified by testing for no autocorrelation in the idiosyncratic disturbances of the equation in differences. This translates to a test for second-order autocorrelation in the first-differenced residuals, which requires $T \geq 5$ (Arellano and Bond 1991). Unfortunately, this replication is limited by the short panel ($T = 3$) available, which precludes testing this assumption with the test statistic for second-order autocorrelation in first-differenced residuals. However, the difference-in-Sargan/Hansen test for the validity of instrument subsets can be employed to investigate the exogeneity of the second-order lag of the dependent variable.

The difference-in-Sargan/Hansen statistic from the model specification in column [2] indicates that the second-order lag of the dependent variable is not a valid instrument (*p-value* = 0.007), suggesting a problem with serial correlation in the residuals. But this result should be interpreted with caution, as the test statistic has been found to over-reject the null hypothesis of exogeneity in the presence of low autocorrelation, with even less power in the presence of more moderate levels of autocorrelation (see Arellano and Bond 1991). Nevertheless, column [4] reports the estimation results from a difference GMM model, which excludes the second-order lag of the dependent variable from the instrument set with robust, Windmeijer-corrected standard errors. This modifies the instrument set for the difference equation to include only the standard instrumental variables specified in the DGHW model from table 5 and the remaining 'GMM-style' instruments employed in the difference GMM models in columns [1] and [2].

In the difference GMM models, the short-term effect of access to all-weather roads is quite large. Columns [1] and [2] assume no serial correlation and indicate roughly a 72–75 percentage point increase in real consumption growth in the short run, which attenuates to a roughly 13.5 percentage point increase at best in the long run. Column [3] reports results under the assumption that there is serial correlation and avoids using the lagged dependent variable as an instrument. This specification indicates that real consumption growth increases by 45.5 percentage points in the short run, with no evidence of a long-run impact from all-weather roads.

The removal of the lagged dependent variable from the instrument set makes this model even more comparable to the model specified by DGHW, which is applied to generate the results in table 5. Comparing the estimates in column [3], it seems that the DGHW model overestimates the short-run impact of all-weather roads. Additionally, the autoregressive coefficient estimate in column [3] is within the credible range, suggesting that the model estimates do not suffer from bias. The original DGHW estimator finds that all-weather roads increase real consumption growth by 16.7 percentage points in the long run. However, in each case the difference GMM results suggest diminishing returns in the long run to this form of investment in capital stocks.

Perhaps the most surprising result is that access to extension is not found to have any impact on the consumption growth trajectory of households after all, in neither the short term nor long term. This is in stark contrast to the result reported in table 5 and the findings for access to extension that DGHW found (that is, a 7.1 percentage point increase in long-run consumption growth). This result holds under the assumption that lagged levels of real consumption are exogenous, as well as the case where serial correlation is assumed to be present. Hence, the results pertaining to access to extension services reported in DGHW do not appear to be robust to changes in the data structure (that is, evenly spaced data), the use alternative instruments via difference-GMM and the Windmeijer correction for the efficient two-step GMM standard errors.[10]

[10] Review of table 2 in DGHW also reveals that extension visits remained sparse over the sample period. No more than 16 per cent of households received extension visits during any year. This could also undermine the ability of regressors to pick up any impact on consumption growth. Thanks are due to an anonymous reviewer for pointing out this issue.

Column [4] reports the result of the same difference GMM estimation strategy with a focus on a different outcome: crop income growth. One would expect that access to extension services would contribute to the growth of crop income for households in the sample. However, it does not appear to make any statistically significant contribution in the short term or long term. On the other hand, access to all-weather roads appears to increase crop income growth in the short run by 33 percentage points, with no long-term effect. It seems reasonable to conclude that increased access to markets after harvesting crops explains this sizable short-run growth contribution. Though it is also of interest to consider the effect of these treatment variables, especially all-weather roads, on off-farm earnings, the number of responses in the sample is too few to provide useful results with the current estimation strategy.[11]

5. Interpretation of the robustness results

At this point, we can synthesise these results to ease interpretation. Recall that the original DGHW estimates are built on assumptions about the evolution of the outcome and key policy variables that are not supported by the data. Henceforth, our objective has been to arrive at a credible empirical model with the least underlying bias given the data structure. An adapted DGHW model was employed thereafter on evenly spaced data to deal with this issue and explicitly consider the role of both short-run and long-run treatment effects. However, this model still appears biased based on the estimates generated for the lagged dependent variable (that is, real consumption). The difference GMM models seek to improve upon the previous biased estimates by including additional and alternative instruments and correcting the standard errors. Table 7 reports the combined results from the key empirical models, which consider the average impact on real consumption growth.

In terms of real money, the first-differenced GMM model indicates that households that gained access to all-weather roads grew real monthly consumption by 111.6 Birr in the short term and 49.5 Birr in the long run (after five years). If we adopt a conservative approach and assume that real consumption is serially correlated, applying the first-differenced GMM model in this context indicates that access to all-weather roads increased real monthly consumption growth by 81.5 Birr in the short term, with no long-term effects. The same first-differenced GMM models indicate that there is no effect on real consumption growth from accessing extension services. Hence, there is a real trade off to be considered within the GoE's portfolio of public investments.

As access to extension services does not appear to be effective in generating positive economic returns for rural households, investments may be diverted towards road infrastructure rather than the current model of extension service delivery. It is easy to

[11] This replication study intentionally avoids extending the analysis to include the system GMM estimator for several reasons. First, the system GMM estimator is best applied to highly persistent data (Blundell and Bond 2000; Bond 2002). The range of credible estimates reported in table 5 suggest otherwise. Second, the initial conditions of the series must be uncorrelated with the fixed effects for system GMM to yield improvements over the first-differenced GMM (Blundell and Bond 1998; Roodman 2009b). This second requirement seems difficult to satisfy in the current setting, as it implies that units of observation must have reached their steady state prior to entering the sample. Given the evidence of high relative growth by some households as chronicled in Dercon *et al*. (2009) (median levels of real consumption grew by 24 per cent between 1994 and 1999) and the existence of low starting points (48 per cent headcount poverty rate in 1994), it seems unlikely that households reached their steady state prior to entering the survey sample.

conceptualise the reasoning behind the differing effects realised from each respective investment channel. Realising economic returns from extension services is an extremely convoluted process. First, extension agents must be well trained and effective at their jobs. Second, farmers must trust agents and the advice that they offer in order for adoption decisions to be made. Even if extension agents can do their jobs well and farmers accept their advice, input delivery systems, which may be out of agents' control, must be able to meet farmers' needs.

Roads, on the other hand, are much simpler, as they ultimately enable rural dwellers to do more of what they do or would do if able. Road investments improve rural dwellers' current movement patterns and even open new routes to previously inaccessible markets, which can generate real benefits as soon as roads are built. However, the economic growth attributable to improved roads is likely to be short lived once the previously unattainable rents have been obtained. If farm productivity, product and labour market conditions remain relatively unchanged, there is no reason to expect households to continue along a positive growth trajectory. Hence, improved roads should be seen as a necessary, but not sufficient, condition for sustained economic growth. This phenomenon is reflected in the current replication results.

Table 7: Effects of public investments on real consumption growth

Empirical model	Public investment	Short run		Long run	
		Percentage-point change in consumption growth rate	Average additional real monthly consumption per adult equivalent, real Birr (1994 = 100)	Percentage-point change in consumption growth rate	Average additional real monthly consumption per adult equivalent, real Birr (1994 = 100)
Original DGHW	Roads	N/A	N/A	16.7	52.7
	Extension services	N/A	N/A	7.1	43.1
Evenly spaced DGHW	Roads	54.7	90.7	–	–
	Extension services	17.0	53	–	–
First-differenced GMM, no serial correlation	Roads	75.6	111.6	13.5	49.5
	Extension services	–	–	–	–
First-differenced GMM, w/ serial correlation	Roads	45.5	81.5	–	–
	Extension services	–	–	–	–

Source: Author's calculations

Note: Statistically insignificant parameter estimates are omitted from this table. Long-run effects should be interpreted as effects from public investments made five years prior to observing the outcome variable. Short-run effects are interpreted as effects from public investments observed in the same period as the observed outcome variable. Additionally, the original DGHW estimator did not consider short-run effects.

6. Conclusion

6.1 Limitations

There are two important limitations to this replication study. The first generates a problem with the amount of data available for the replication, and the second is a product of the subsequent data shortage. Both issues create challenges with the application of the dynamic panel GMM estimator employed to evaluate the robustness of the results found in DGHW.

Given the public release of the 2009 round of the ERHS data used in DGHW, this replication study also attempted to incorporate the new round of data into the analysis. Although variable construction was fairly straightforward, there were several problems with the matching of observations from the 1994–2004 data to the 2009 data. This was largely due to the change in the coding of the clustered identifiers. For instance, in 2004, woredas found in 2009 data are coded as regions, and regions have been redefined in 2009 to reflect the geographic boundaries of the major national regions in Ethiopia (that is, SNNPR, Tigray, Amhara and Oromia).[12] Additionally, some of the woredas are coded differently, with different names and identifiers. These identifiers are essential to constructing unique household identifiers within the publicly available data set. To construct the identifiers, I follow the procedure described in the data description document that accompanies the ERHS data files.

The size of the sample found in the data files for round 7 (2009) is 99.48 per cent of the sample size in the previous round (2004). I attempted to work through the coding challenges as best as possible by meticulously searching through the data files to identify matches and recode some of the identifiers in the data. However, despite my best efforts, some discrepancies persisted. Due to the discrepancies with the coding and identifiers, merging the data across these rounds reduces the sample size by at least 17.5 per cent. This is a nontrivial amount of unmatched data, which makes the analysis of the 2009 round less appealing and increases the potential that the analysis will be unreliable or incomparable to the original study.

As previously noted, the validity of the second-order lag of the dependent variable as an instrument relies strongly on the absence of autocorrelation in the residuals on the consumption growth model. Due to the short panel available to this replication, it is not possible to directly test for second-order serial correlation in the first-differenced residuals.[13] Although prone to over-rejection of the null in the presence of serial correlation, the difference-in-Sargan/Hansen test gives insight into the validity of this instrument and suggests that the difference GMM specification, which excludes the lagged dependent variable from the instrument set, may be more reliable.

[12] Woredas are districts or the third-level administrative divisions of Ethiopia. They are composed of a number of wards (kebele) or neighborhood associations, which are the smallest unit of local government in Ethiopia.

[13] If the 2009 data could have been implemented, then we would have $T = 4$, which would make the third-order lag of the dependent variable a valid instrument even if there existed second-order serial correlation in the first-differenced residuals.

6.2 Summary and final remarks

This replication study reviews a popular study of the effect of access to technology and access to capital stocks in the form of all-weather roads on a sample of rural Ethiopian households. The original findings suggest that the long-term impact of these public investments contributed to considerable and positive gains in consumption growth rates within the sample. The pure replication supports the original authors' findings under the central assumption that household consumption levels, capital stock and access to technology change slowly over time. However, this replication formally checks the robustness of these assumptions as well as the estimation strategy employed in the original study with an alternative dynamic panel GMM estimator, a modified set of instrumental variables and evenly spaced panel data. The results of the replication suggest that the qualitative results previously put forward may be only partially reflected by the data, with the need for a reinterpretation of the implications.

First, the original findings relied upon strong assumptions about the evolution of the data for key variables. This replication has shown that those underlying assumptions do not hold within the data. Though these assumptions allow the use of more data across survey rounds, the irregular frequency of the data collection creates a bias and invalid inference due to a Mixed Data Sampling problem (Andreou *et al*. 2010). Second, the specification in the original paper focuses only on the long-term effect of public investments, without the inclusion of important short-run variables of interest where they are theoretically relevant and empirically possible. This modifies the overall interpretation of the original findings by placing the reported impact in a long-run context only, but it also exposes the original estimation to omitted variables bias. Additionally, some of the instruments deployed in the original study are shown to be invalid when the same model specification is placed in a dynamic panel context with evenly spaced data.

Furthermore, the estimator deployed in the original study appears to generate biased estimates of standard errors in the context of a dynamic panel. The original authors apply an efficient two-step GMM estimator, which controls for fixed effects and is robust to heteroskedasticity but is also known to yield standard errors with a downwards bias (Arellano and Bond 1991; Windmeijer 2005). This seems to have the greatest implications for the result on access to extension services. This replication shows that the conclusion about the impact of access to extension services is not robust to a similar dynamic GMM model, which employs a modified instrument set and uses Windmeijer-corrected standard errors to draw inferences.

The diagnostics analysis in this replication suggest that there may be serial correlation present, which necessitates the use of the difference GMM estimator without the use of the lagged dependent variable as an instrument. Under this specification, the estimators (that is, DGHW and difference GMM) are easier to compare in terms of their respective instrument sets, with key differences in the treatment of standard errors. Nonetheless, after adjusting the interpretation of the original results, part of the story remains.

Gaining access to all-weather roads leads to significant and positive short-term effects on real household consumption growth and crop income growth. A conservative estimate suggests an 81.5 Birr increase in real monthly consumption per adult-equivalent unit. This short-run effect is much larger than the long-run effect reported by DGHW. Real crop income growth is also found to increase significantly in the short run for households with access to all-weather roads. However, these growth effects attributable to road development found in this replication study are found to diminish completely in the long run (that is, after five years). As reflected by the replication study results, this implies that the largest benefits to household consumption and crop income growth due to rural road development accrue in the short run. This is an important policy finding that can help with time-sensitive targeting of other interventions or public programmes designed to uplift the rural poor. Road development does not appear to be sufficient to sustain economic growth for the poor.

However, the results found in the original study concerned with the impact of access to extension services and agricultural technologies do not hold up as well to the robustness checks. This replication study shows that there is no statistically meaningful contribution towards household consumption growth in this sample as a result of receiving a visit from an extension agent. Similarly, there is no effect from extension visits on crop income growth. This may call into question the efficacy of the prevailing extension services during the time of the sample.

During the study period, Ethiopian extension agents functioned as a source of credit and inputs rather than as a source of advice on optimal use of inputs and land management (Spielman *et al*. 2011). Access to technology without knowledge of how best to use it is not sufficient to improve crop performance or other indicators of rural household welfare. Hence, the emphasis of extension agents as suppliers of inputs and credit seems to have resulted in ineffective knowledge transfer. Furthermore, Ethiopian extension agents have been documented as having little practical experience and having poor communication skills (Belay and Abebaw, 2004). Mogues *et al*. (2009) report that while extension officers may make contact with farmers, the farmers typically do not adopt extension advice. This makes the GoE’s strategy of expanding the public extension service programme since 2008 to reach more farming areas throughout the country problematic.[14] The evidence points to a case where the advice of extension officers with poor practical skills falls on deaf ears. Given the reported inefficiency of the Ethiopian agricultural extension service system, it is not surprising that the replication results do not show evidence of an impact from extension services. On the other hand, perhaps the value of extension services could be better assessed by measuring crop productivity gains. However, this may be a case for investigation beyond the scope of the current replication.

[14] See http://hornaffairs.com/en/2012/06/05/extension-service-the-driver-of-ethiopias-agricultural-revolution-abraham-dereje/

References

Andersen, TG and Sørensen, BE, 1996. GMM estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Business & Economic Statistics*, 14(3), pp.328–352.

Andreou, E, Ghysels, E and Kourtellos, A, 2010. Regression models with mixed sampling frequencies. *Journal of Econometrics* 158(2), pp.246–261.

Arellano, M and Bond, S, 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58(2), pp.277–297.

Baum, C, Schaffer, M and Stillman, S, 2010. IVREG2: Stata module for extended instrumental variables/2sls, gmm and ac/hac, liml and k-class regression. Available at: <http://ideas.repec.org/c/boc/bocode/s425401.html> [Accessed 6 January 2015]

Belay, K and Abebaw, D, 2004. Challenges facing agricultural extension agents: a case study from south-western Ethiopia. *African Development Review*, 16(1), pp.139–168.

BenYishay, A and Tunstall, R, 2011. Impact evaluation of infrastructure investments: the experience of the Millennium Challenge Corporation. *Journal of Development Effectiveness*. Available at: <http://www.tandfonline.com/doi/abs/10.1080/19439342.2010.545892> [Accessed 6 January 2015]

Blundell, R and Bond, S, 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), pp.115–143.

Blundell, R and Bond, S, 2000. GMM estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19(3), pp.321–340.

Bond, S, 2002. Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal*, 1(2), pp.141–162.

Bond, S, Hoeffler, A and Temple, J, 2001. *GMM estimation of empirical growth models*. CEPR Discussion Paper no. 3048. London: Centre for Economic Policy Research.

Bowsher, CG, 2002. On testing overidentifying restrictions in dynamic panel data models. *Economics Letters*, 77(2), pp.211–220.

Dalgaard, CJ, Hansen, H and Tarp, F, 2004. On the empirics of foreign aid and growth. *Economic Journal*, 114(496), pp.F191–F216.

Dercon, S, Gilligan, DO, Hoddinott, J and Woldehanna, T, 2009. The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages. *American Journal of Agricultural Economics*, 91(4), pp.1,007–1,021.

Dercon, S and Zeitlin, A, 2009. Rethinking agriculture and growth in Ethiopia: a conceptual discussion. *Department for International Development, United Kingdom*. Available at: <http://users.ox.ac.uk/~econstd/ethiopia%20paper%202_v2.pdf> [Accessed 6 January 2015]

Dorosh, P and Schmidt, E, 2010. The rural-urban transformation in Ethiopia. *International Food Policy Research Institute*. Available at: <http:// http://www.ifpri.org/sites/default/files/publications/esspwp013.pdf> [Accessed 6 January 2015]

Hansen, LP, 1982. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pp.1,029–1,054.

Holtz-Eakin, D, Newey, W and Rosen, HS, 1988. Estimating vector autoregressions with panel data. *Econometrica: Journal of the Econometric Society*, pp.1,371–1,395.

Khandker, SR and Koolwal, GB, 2011. Estimating the long-term impacts of rural roads: a dynamic panel approach. *World Bank Policy Research Paper Series, no. 5867.* Washington, DC: World Bank.

Maffioli, A, Ubfal, D, Baré, GV and Cerdán-Infantes, P, 2011. Extension services, product quality and yields: the case of grapes in Argentina. *Agricultural Economics* 42(6), pp.727–734.

Mardia, KV, Kent, J and Bibby, J, 1976. *Multivariate Analysis*. London: Academic Press.

Minoiu, C and Reddy, SG, 2010. Development aid and economic growth: a positive long-run relation. *Quarterly Review of Economics and Finance*, 50(1), pp.27–39.

Mogues, T, 2011. The Bang for the Birr: Public Expenditures and Rural Welfare in Ethiopia. *Journal of Development Studies*, 47(5), pp.735–752.

Mogues, T, Cohen, MJ, Birner, R, Lemma, M, Randriamamonjy, J, Tadesse, F and Paulos, Z, 2009. *Agricultural extension in Ethiopia through a gender and governance lens*. International Food Policy Research Institute (IFPRI).

Roodman, D, 2009a. How to do xtabond2: An introduction to difference and system gmm in stata*. Stata Journal*, 9(1), pp.86–136.

Roodman, D, 2009b. A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71(1), pp.135–158.

Spielman, DJ, Kelemwork, D and Alemu, D, 2011. *Seed, fertilizer, and agricultural extension in Ethiopia*. ESSP II Working Paper 20.

Windmeijer, F, 2005. A finite sample correction for the variance of linear efficient two-step gmm estimators. *Journal of Econometrics*, 126(1), pp.25–51.

Replication Paper Series

International Initiative for Impact Evaluation
1625 Massachusetts Ave., NW
Suite 450
Washington, DC 20036
USA

replication@3ieimpact.org
Tel: +1 202 629 3939

www.3ieimpact.org