Valuing Water Purification by Forests: An Analysis of Malaysian Panel Data

Jeffrey R. Vincent, Ismariah Ahmad, Norliyana Adnan, Walter B. Burwell, Subhrendu K. Pattanayak, Jie-Sheng Tan-Soo, et al.

Environmental and Resource Economics
The Official Journal of the European Association of Environmental and Resource Economists

ISSN 0924-6460

Environ Resource Econ
DOI 10.1007/s10640-015-9934-9

Springer

Environ Resource Econ
DOI 10.1007/s10640-015-9934-9

Valuing Water Purification by Forests: An Analysis of Malaysian Panel Data

Jeffrey R. Vincent[1] · Ismariah Ahmad[2] · Norliyana Adnan[2] · Walter B. Burwell III[1] · Subhrendu K. Pattanayak[1] · Jie-Sheng Tan-Soo[1] · Kyle Thomas[1]

Accepted: 11 June 2015

Abstract Water purification might be the most frequently invoked example of an economically valuable ecosystem service, yet the impacts of upstream land use on downstream municipal water treatment costs remain poorly understood. This is especially true in developing countries, where rates of deforestation are highest and cost-effective expansion of safe water supplies is needed the most. We present the first econometric study to estimate directly the effect of tropical forests on water treatment cost. We exploit a rich panel dataset from Malaysia, which enables us to control for a wide range of potentially confounding factors. We find significant, robust evidence that protecting both virgin and logged forests against conversion to nonforest land uses reduced water treatment costs, with protection of virgin forests reducing costs more. The marginal value of this water purification service varied greatly across treatment plants, thus implying that the service offered a stronger rationale for forest protection in some locations than others. On average, the service value was large relative to treatment plants' expenditures on priced inputs, but it was very small compared to producer surpluses for competing land uses. For various reasons, however, the latter comparison exaggerates the shortfall between the benefits and the costs of enhancing water purification by protecting forests. Moreover, forest protection decisions that appear to be economically unjustified when only water purification is considered might be justified when a broader range of services is taken into account.

Keywords Ecosystem service · Water purification · Forest · Malaysia · Valuation

✉ Jeffrey R. Vincent
Jeff.Vincent@duke.edu

1 Nicholas School of the Environment, Duke University, Durham, NC 27708, USA

2 Forest Research Institute Malaysia, 52109 Kepong, Selangor, Malaysia

Published online: 24 June 2015

1 Introduction

Water purification might be the most frequently invoked example of an economically valuable service provided by ecosystems to human society. The concept of this service is straightforward: water that runs off or seeps through forests and other natural ecosystems tends to be less contaminated than water discharged by agricultural, urban, or industrial landscapes; hence, it requires less costly treatment before it is fit to drink. Water with a high concentration of silt and other suspended solids is especially costly to treat, as it requires a multi-stage treatment process (coagulation, sedimentation, filtration, disinfection; http://water.epa.gov/learn/kids/drinkingwater/watertreatmentplant_index.cfm). An econometric literature that dates back to the 1980s confirms that the operating costs of treatment plants are lower when source water is cleaner (Forster et al. 1987; Moore and McCarl 1987; Holmes 1988; Dearmont et al. 1998; Murray and Forster 2001; Piper 2003). A separate case-study literature has examined investments in watershed protection programs by individual water utilities. It reports that, by improving source water quality, the programs allow utilities to use simpler treatment processes and thereby avoid large capital costs (Chichilnisky and Heal 1998; Dudley and Stolton 2003; Barten and Ernst 2004; Postel and Thompson 2005; Sklenar et al. 2012; Alcott et al. 2013; Gartner et al. 2013; Spiller et al. 2013). The best-known case is New York City, which reportedly avoided the multi-billion-dollar cost of a new filtration plant by restoring ecosystems in its Catskills watershed (Chichilnisky and Heal 1998).

Despite this prior work, the impacts of land use on water treatment costs remain poorly understood. Hydrological research offers abundant evidence that runoff from forests tends to have lower concentrations of suspended solids than runoff from other land uses (Dunne and Leopold 1978; Hewlett 1982; Bruijnzeel 2004; Carlson et al. 2014), but the effects of forests on water quality vary in large watersheds with multiple land uses (Dissmeyer 2000; Bruijnzeel 2004; Hurley and Mazumder 2013). Econometric studies have largely ignored the effects of land use, as opposed to water quality, on treatment costs.[1] An exception is a study by Ernst (2004; see also Ernst et al. 2004), which found that "for every 10 % increase in forest cover in the source area, treatment and chemical costs *decreased* approximately 20 %" (p. 21; italics in the original). When Freeman et al. (2008) augmented Ernst's data with data from additional treatment plants, however, the effect became smaller and statistically less significant. The case studies on watershed protection programs directly address the effects of land use, but they often report simulated, prospective savings, not actual savings. Sklenar et al. (2012, p. 12) report that water utilities are reluctant to invest in these programs because information on the programs' benefits and costs is inadequate. The role of watershed management in reducing capital costs has been debated even in the Catskills case (Sagoff 2002; Kenny 2006).

Here, we present an econometric study that directly analyzes the effect of forest cover on water treatment cost. We extend the literature in two important ways. First, we control more carefully than prior studies for factors that could confound identification of this effect. Our dataset is a panel with monthly observations on 41 treatment plants during 1994–2007. It is much larger than the datasets analyzed by the econometric studies cited above, and this enables us to use fixed effects to control for unobserved characteristics of treatment plants, their catchments, and time periods. The studies by Ernst (2004) and Freeman et al. (2008) analyzed cross-sectional datasets of just 25 and 40 observations, respectively, and neither one controlled for variables other than land use. Of the econometric studies that investigated the effects of water quality, one analyzed a single treatment plant (Moore and McCarl 1987)

[1] The process-based model InVEST is increasingly being used to predict water treatment costs under different land-use scenarios (Conte et al. 2011).

and three analyzed no more than a dozen (Forster et al. 1987; Dearmont et al. 1998; Murray and Forster 2001). Only two analyzed panels (Forster et al. 1987; Dearmont et al. 1998), and both panels were very short (2 and 3 years, respectively); moreover, neither study exploited the panel structure to control for unobserved effects.

Second, to the best of our knowledge, our study is the first econometric analysis on the effect of forest cover on water treatment costs in the developing world. Most of the literature cited above, including all of the econometric studies, refers to developed countries, mainly the US. Our study site is the Malaysian state of Perak, which is an attractive site due to the availability of not only a large panel dataset on water treatment plants but also forest cover data that distinguish virgin (undisturbed) forests from logged forests. In addition, treatment plants in Perak use the concentration of suspended solids as the "main source of guidance on chemical use" (Mohd Nordin et al. 2000).[2] This creates a plausible mechanism relating forests to treatment costs, if forests indeed reduce suspended solids in source water.

Two previous empirical studies have investigated environmental aspects of water treatment costs in the tropics, and both analyzed very small samples: monthly data for a single year for one treatment plant in Thailand (Sthiannopkao et al. 2007) and six plants in Malaysia (Abdul Rahim and Mohd Shahwahid 2011). Both reported that operating costs were lower in months when source water was cleaner, but neither one provided empirical evidence on the effect of forest cover on either water quality or treatment cost. A recent global study predicted improvements in water quality and reductions in treatment costs that could result from investments in watershed management programs by 534 cities, including many in the developing world (McDonald and Shemie 2014), but it estimated the effect of water quality on treatment cost using cross-sectional data from only US cities and a simple, uncontrolled regression, and it simulated the effect of watershed management on water quality.

Developing a better understanding of the value of water purification by forests is especially important in developing countries because rates of deforestation and forest degradation, and thus threats to this service, are much higher than in developed countries (FAO 2010). Attempts to reduce these threats, including via payment programs for watershed services (Bennett et al. 2013), have been expanding rapidly in developing countries, with little empirical evidence on the quantity or value of the services supplied by the conservation interventions that the programs fund (Ferraro et al. 2012). Developing a better understanding of the service value in developing countries is also important in view of the much higher percentage of the population without access to safe drinking water.[3] Millennium Development Goal 7, "Ensure Environmental Sustainability," includes as a target to "Halve, by 2015, the proportion of people without sustainable access to safe drinking water and basic sanitation" (http://www.unmillenniumproject.org/goals/gti.htm). Water-borne and water-related diseases continue to loom large in the global burden of disease, especially for children in the developing world. Given that the cost of improved municipal water supply in developing countries generally exceeds households' willingness to pay for it (Whittington 2010), forest protection might

[2] Treatment plants in Perak measure the concentration of suspended solids in source water as it enters the plant, and they adjust the amount of alum, the chief chemical used to remove suspended solids, accordingly. Alum reduces the pH of treated water, thus requiring the addition of lime before the water is distributed and causing a knock-on increase in treatment costs (Mohd Nordin et al. 2000). The third major treatment chemical is chlorine, a disinfectant, but treatment plants in Perak add it as a standard dose that does not vary with source water quality (personal communication, Perak Water Board, February 10, 2014).

[3] The proportion of the population with access to piped water or another improved water source is higher in Malaysia than in most developing countries: 88.2 % in 1990 and 99.6 % in 2010, according to the World Bank's World Development Indicators database (http://data.worldbank.org/data-catalog/world-development-indicators). Currently, 98 % of Perak households have a piped water connection (personal communication, Perak Water Board, May 26, 2015).

be able to enhance progress toward this target if forests do indeed have a substantial cost-reducing effect.

The paper is organized as follows. The next section presents our panel model and discusses associated econometric issues, construction of the variables in the model, and data sources. The subsequent section presents our econometric results, which provide strong evidence that protecting virgin forests against conversion to nonforest land uses reduced treatment costs in Perak. Protecting logged forests against conversion also reduced costs, but by a smaller amount. We use the econometric results to estimate the marginal value of the water purification service supplied by virgin forests, and we find that the value varies greatly across treatment plants. The final section discusses several issues that must be considered when comparing estimates of this heterogeneous value to the marginal cost of forest protection.

2 Materials and Methods

2.1 Econometric Approach

The objective of our econometric analysis was to identify the marginal effect of land use—specifically, forest cover—on water treatment cost. The analysis was guided by the theory of using cost functions to value environmental inputs (McConnell and Bockstael 2005, p. 647; Vincent 2011, pp. 46–54; Freeman et al. 2014, pp. 240–244). It was strongly influenced by two features of our data: the data were a panel, which enabled us to use fixed effects to control for unobserved factors that might confound the effect of land use on cost; and the cost data referred to operating cost, which meant we were analyzing a short-run effect.

A generic short-run cost function for a firm that uses one or more unpriced environmental inputs includes four types of variables: (1) the firm's output level, (2) prices paid by the firm for labor and other nonenvironmental inputs, (3) the quantity of capital and other fixed factors used by the firm, and (4) the quantity of environmental inputs used by it (McConnell and Bockstael 2005, p. 629; Vincent 2011, pp. 46, 67–69). This led us to specify the following fixed-effects regression model (Wooldridge 2010, Ch. 10–11):[4]

$$\ln\left(C_{it}\right)=\ln\left(\mathbf{L}_{iy}\right)\boldsymbol{\beta}+\alpha\ln\left(Q_{it}\right)+\gamma_1\ln\left(R_{it}\right)+\gamma_2\left(\ln\left(R_{it}\right)\right)^2+c_i+\theta_y+\theta_m+u_{it}. \quad (1)$$

C_{it} is the operating cost of treatment plant (TP) i in time period t, which is a given month m of a given year y.[5] $\mathbf{L}_{iy}$ is a matrix of land-use variables, which varied by year but not months within a year. Q_{it} is treated water volume (i.e., the quantity produced), and R_{it} is rainfall; both varied by year and month. Land use and rainfall refer to a TP's catchment. $\boldsymbol{\beta}$, α, γ_1, and γ_2 are coefficients to be estimated; c_i, θ_y, and θ_m are fixed effects for TPs, years, and months, respectively; and u_{it} is the error term.

Comparing Eq. (1) to a generic cost function, water volume (Q_{it}) obviously represents the first variable type (output level). The year effects (θ_y) control for the second type (input

[4] This double-log specification for treatment cost and water volume is common in the econometric literature on water treatment (Forster et al. 1987; Moore and McCarl 1987; Holmes 1988; Murray and Forster 2001; Piper 2003). We confirmed the appropriateness of the log-transform of treatment cost by estimating a Box-Cox model, which favored that specification over a linear specification: the estimated theta parameter in a left-hand-side Box-Cox model was 0.117, which is much closer to 0 than to 1. We confirmed the appropriateness of the log-transforms of the other explanatory variables by comparing the overall fit of models containing the transformed variables to ones containing untransformed variables.

[5] Some prior studies have used unit cost (C/Q) as the dependent variable (e.g., Dearmont et al. 1998; Piper 2003; Ernst 2004). Our double-log specification encompasses this as a special case ($\alpha = 1$).

prices). Three inputs—labor, energy, and treatment chemicals—account for nearly all of the operating expenditures of TPs in Perak; the fourth operating expenditure category, maintenance, equals zero in most months. Prices of these inputs vary over time but not across the TPs in the sample, which are all operated by the same government agency, the Perak Water Board. This agency has a civil-service pay scale, purchases electricity from the national grid, and pays for the chemicals used by the individual TPs. The TP effects (c_i) control for the third variable type (capital and other fixed factors). For TPs that were upgraded during the sample period, we allowed the TP effects to differ before and after the upgrades.

Finally, the land-use variables ($\mathbf{L}_{iy}$) and rainfall (R_{it}) represent the fourth variable type (environmental inputs). The effect of land use is our main interest, but controlling for rainfall is important because rainfall might affect changes in land use, for example by impeding the burning of woody debris when forests are converted to nonforest uses. Controlling for rainfall thus reduces the risk of biased estimation of the effect of land use on treatment cost. The fixed effects similarly control for the potentially confounding effects of a wide range of unobserved variables, both environmental and nonenvironmental: the TP effects (c_i) control for time-invariant TP characteristics, including catchment area, topography, soils, and geology; the year effects (θ_y) control for TP-invariant annual characteristics, including climate trends and changes in state and national environmental and natural resource policies; and the month effects (θ_m) control for TP-invariant monthly characteristics, such as seasonality. Given the inclusion of the TP effects, we identified the effect of land use on treatment cost purely from the variation in the data over time ("within variation"), not the variation between TPs.

An individual variable L in $\mathbf{L}$ was defined as the share of the surface area of a TP's catchment that was in a particular land use in a particular year. Denoting forest area by F and catchment area by A, $L = F/A$. The Malaysian forest inventory data enabled us to distinguish three land uses: virgin forest, logged forest, and nonforest. Because the shares summed to unity across these categories, we excluded one category—the nonforest share—which became the reference land use against which the effects of the others were measured. So, a negative coefficient β on a particular forest variable in L indicated that avoiding conversion of that type of forest to nonforest land use reduced treatment cost, while a positive coefficient indicated that avoided conversion raised cost.

Because virgin forest area equaled zero in some catchments, we used $1 + F/A$ instead of F/A as the forest cover variable for both virgin and logged forests. Without this modification, the elasticity of treatment cost C with respect to F would equal the estimated coefficient $\hat{\beta}$; with the modification, the elasticity is given by

$$\hat{\beta}\frac{F/A}{1 + F/A}. \tag{2}$$

The marginal effect of F on C is given by partially differentiating expected treatment cost from Eq. (1) with respect to F, $\partial \mathrm{E}\left(C \left| \mathbf{L}, Q, R, c_i, \theta_y, \theta_m\right.\right)/\partial F$, which yields

$$\frac{\hat{\beta} \cdot \mathrm{E}\left(C \left| \mathbf{L}, Q, R, c_i, \theta_y, \theta_m\right.\right)/A}{1 + F/A}, \tag{3}$$

where $\mathrm{E}\left(C \left| \mathbf{L}, Q, R, c_i, \theta_y, \theta_m\right.\right)$ is treatment cost predicted by the regression.

Assuming $\hat{\beta} < 0$, the negative of Eq. (3) provides an estimate of the value of the water purification service supplied by a marginal hectare of forest. It represents a conservative estimate (i.e., an underestimate), for two reasons. First, given that C refers to just operating costs, Eq. (3) omits potential reductions in capital costs of the sorts emphasized in the case-study literature on water treatment cost and land use (Holmes 1988, p. 361). This could be a substantial omission: Mohd. Akbar and Rusnah (2004, p. 7) list six TPs in Malaysia that were "forced … to be abandoned at a substantial loss," and another five that "had to be upgraded technically incurring huge costs in order to maintain their utility," due to degradation of source water quality caused by "extensive logging and land clearing." None of the listed TPs was in Perak, but as noted below several TPs in our sample closed during the sample period, and a larger number were upgraded.[6] If these actions were taken in response to degraded source water quality, then we have underestimated the value of the water purification service, as cleaner source water would have extended the lifetime of the TPs and enabled them to avoid the capital costs of upgrading. Similarly, cleaner source water might have reduced the original capital costs of some of the TPs.

The second reason is related to the fact that, as a partial derivative, Eq. (3) holds water volume, Q, constant. If TPs are able to treat more water when their source water is cleaner, for example because production is less constrained by retention time in the plants' sedimentation tanks (Mohd Nordin et al. 2000), then the net return on the additional treated water is also part of the marginal value of the water purification service. Eq. (3) omits the value of this "extensive margin" effect, however. Treating output quantity as fixed is a well-known cause of undervaluation when cost functions are used to value environmental inputs (McConnell and Bockstael 2005, p. 647; Vincent 2011, pp. 63).[7]

This second reason raises a further econometric issue: if TP managers can control the volume of water they treat, at least up to a TP's capacity, then Q is potentially endogenous. This too is a well-known issue with cost functions (Vincent 2011, pp. 59–61), including cost functions for water treatment (Piper 2003, pp. 6–7). In Sect. 3 , we will present evidence that Q was not endogenous in our sample and that a presumptive instrumental variables (IV) model yields results virtually identical to those from an OLS model.

Prior research enables us to sign the expected effects of water volume and rainfall in Eq. (1). (We have no priors about the signs of the fixed effects, which control for multiple unobserved factors.) We expect treatment cost to increase less than proportionally with water volume (Forster et al. 1987; Holmes 1988; Dearmont et al. 1998; Murray and Forster 2001; Piper 2003), thus bounding $\hat{\alpha}$ by 0 and 1. Regarding rainfall, much literature indicates that rainfall increases sediment loads in rivers in tropical regions, as it erodes soil and transports sediment as it runs off the land (Dunne 1979; Krishnaswamy et al. 2001; Bruijnzeel 2004; Abdul Rahim and Zulkifli 2004). There is also evidence from Thailand (Sthiannopkao et al. 2007) and, outside the tropics, Texas (Dearmont et al. 1998) that treatment cost is higher during wetter periods. Holmes (1988, p. 362) observes, however, that dilution can cause the concentration of contaminants in source water to be lower, and thus treatment cost to be lower too, when streamflow is greater. Given that rainfall increases streamflow, this dilution effect could similarly cause treatment cost to be lower when rainfall is higher. To allow this effect, we modeled the effect of rainfall as quadratic. We hypothesized that the dilution effect would dominate the soil erosion effect and reduce cost at lower, less erosive rainfall levels ($\gamma_1 < 0$) but not at higher, more erosive rainfall levels ($\gamma_2 > 0$).

[6] In addition, data were incomplete for some TPs: 32 % of the TP-years had fewer than 12 months of data, with 19 % having fewer than 11 months. It is possible that some of the missing observations resulted from outages or failures caused by poor source water quality. We doubt that this explains most of the missing observations, however, which were typically missing for all or nearly all of the TPs in the state or a particular region of the state in a given month. This pattern implies problems with data recording and storage, not temporary closures of individual TPs.

[7] This problem can be avoided by estimating a profit function instead of a cost function. See Pattanayak and Kramer (2001) for an application to a different watershed service, drought mitigation. We were unable to estimate a profit function because the Perak Water Board did not provide data on revenue from water sales.

A final econometric issue concerns the standard errors of the coefficients. We used "sandwich" estimates (Huber 1967; White 1980), which are robust to heteroskedasticity. Conventional robust standard errors (and OLS standard errors too) can underestimate true standard errors and exaggerate significance when an explanatory variable varies at a higher (more aggregate) level than the dependent variable does (Moulton 1986). This "Moulton problem," or pseudoreplication as it is known to ecologists, was a risk in our data, given that land use was annual while treatment cost was monthly. It can be addressed by clustering standard errors at the higher level (Angrist and Pischke 2009, Ch. 8), which in our case implied clustering them by year for each TP. To be additionally conservative, we clustered at an even higher level—by TP—which presumptively corrected for serial correlation within and between years (Zeger and Liang 1986).

2.2 Variable Definitions and Data Sources

We obtained TP-level data on treatment cost and treated water volume from the Perak Water Board, a state-owned utility that operated 46 TPs during 1994–2008. These are most of the TPs in the state.[8] We excluded four TPs that drew water from a source other than run-of-the-river (e.g., a reservoir)[9] and one without reliable information on the location of its water intake.[10] The final sample thus included 41 TPs (Fig. 1). The fact that we dropped only one qualified TP from our sample makes our sample more complete than many in the literature; for example, incomplete information forced Holmes (1988) and Piper (2003) to drop more than half of the TPs or water utilities in their US samples. Most (33) of the TPs in our sample were established before 1994 and operated the entire period; of the remaining eight, four were established before 1994 and closed before 2008,[11] while four were established after 1994 and operated through 2008.[12] The panel was thus unbalanced. Eleven of the TPs were upgraded during 1994–2008 (Lembaga Air Perak n.d., pp. 50–51).[13]

The TPs in the sample used a standard treatment process: raw water was extracted from a river, screened to remove debris, aerated to remove gases and oxidize impurities, mixed with alum to coagulate suspended solids, passed through sedimentation tanks and then sand filters to remove the coagulated solids ("floc"), and finally dosed with lime to adjust pH and with chlorine to kill pathogens. The TPs varied mainly in terms of capacity, not technology. Malaysia's drinking water quality standards[14] match those in the US[15] for many parameters (e.g., zero total coliforms and E. coli; 250 mg/l chloride) but are somewhat less stringent for others (notably, 1000 mg/l total dissolved solids instead of 250 mg/l). Water treatment costs

[8] A private company, the Malaysian Utilities Corporation, operates two additional TPs in Perak (Sultan Idris Shah II, Ulu Kinta). It operated a third until 2001 (Pekan Parit). A second company, GSL Water, operates two others (Gunung Semanggol, Taiping Headworks). We excluded these private TPs from the analysis due to their different management and because the companies did not report plant-level data in all time periods.

[9] Air Kuning, Gerik V, Jalan Baru, and Pulau Banding. The exclusion of TPs that used reservoirs probably biased our estimate of the effect of forests on treatment costs downward, as we have ignored dredging costs.

[10] Selama, which is one of the newest TPs (established in 2004).

[11] Grik, Sg. Jelintoh, and Ulu Soh, which closed in 2000, and Felda Bersia, which closed in 2001.

[12] BB Seri Iskandar (opened in 2001), Sg. Tapah (2002), Tg. Malim (2003), and Hilir Perak (2006).

[13] Perlop (1995), Terong (1999), Padang Rengas (2001), Teluk Kepayang (2001), Kota Lama Kiri (2002), Sg. Kampar (2004), Kg. Paloh (2005), Pengkalan Hulu (2005), Jelai (2007), Sumpitan (2007), and Manong (2008).

[14] A complete list is available at: http://kmam.moh.gov.my/public-user/drinking-water-quality-standard.html.

[15] A complete list is available at: http://water.epa.gov/drink/contaminants/#List.

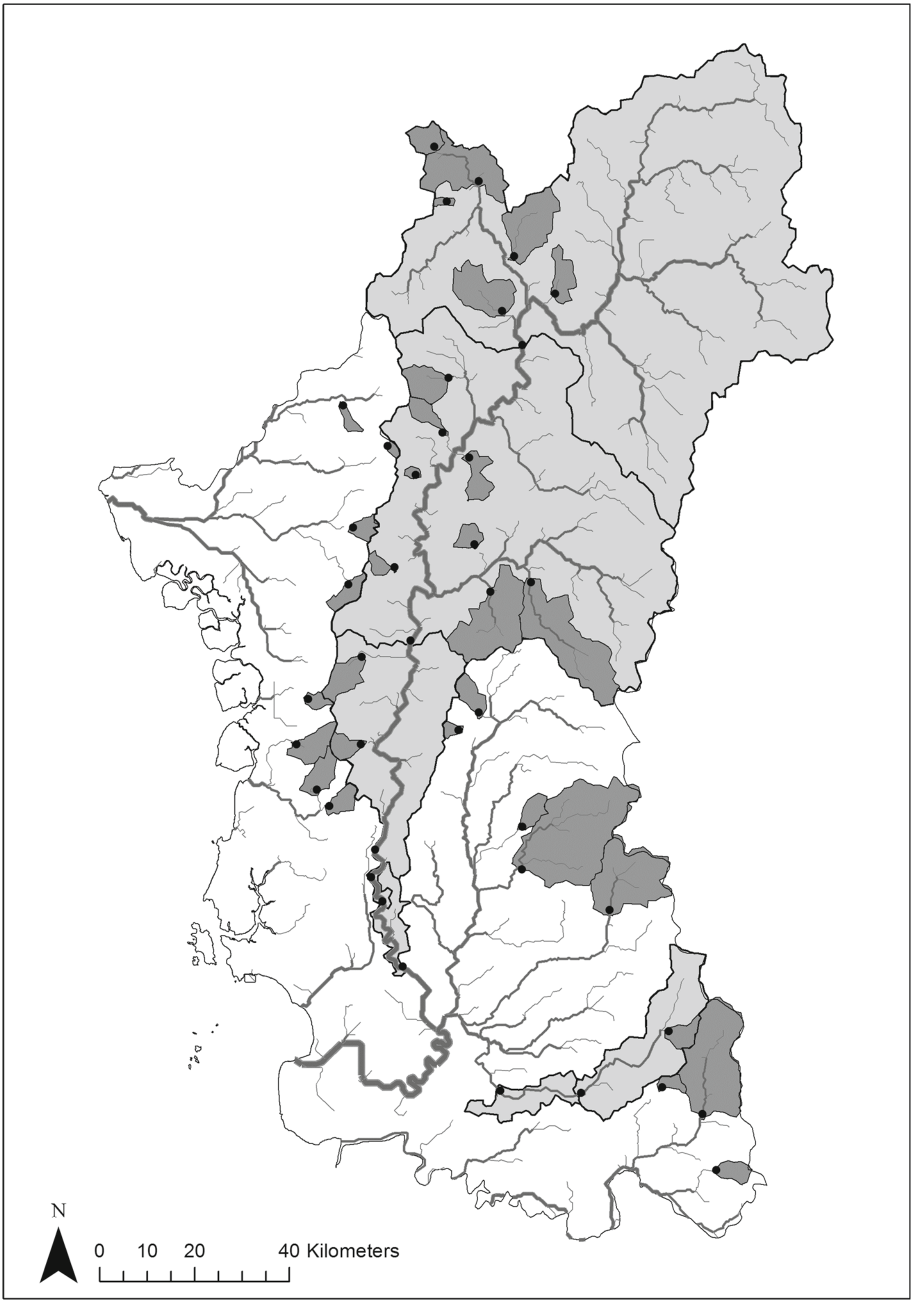

Fig. 1 Map of Perak showing major rivers and streams and the locations of water intakes (*black dots*) and catchments (*gray shading*, *black boundaries*) for the 41 TPs in the regression sample. *Light gray*: larger catchments, *dark gray*: smaller catchments. The Perak River is the large river that flows north–south through the state. The state's western border is the Strait of Malacca; other borders are with Thailand (north) and other Malaysian states (northwest, east, south)

and thus the potential benefits of cleaner source water would have been higher if the TPs had been required to meet the more stringent US standards. This is another reason our results underestimate the value of the water purification service in Perak.

The Perak Water Board provided summary monthly data on the TPs in multiple spreadsheets that varied in format and coverage (number of TPs, number of months). We extracted data on total operating cost and water volume.[16] Total operating cost was expressed in the Malaysian currency, the ringgit (RM; 2005 exchange rate: 3.79 RM/US$). We used the Malaysian GDP deflator to convert it to 2005 price levels. Treated water volume was expressed in cubic meters. The number of observations with data on both cost and water volume was 4246.[17]

We obtained land-use data from the Forestry Department Peninsular Malaysia, which has conducted National Forest Inventories (NFIs) approximately every ten years since 1971. The Department granted us restricted access to GIS layers from NFI 3 (1992) and NFI 4 (2004), which enabled us to classify catchment area into virgin forest, logged forest, and nonforest land use. The forests in the catchments are tropical rainforests located mostly on hilly or mountainous terrain. The coastal plains of the Peninsula were once covered by lowland rainforests, but most of these forests were already converted to nonforest land uses, mostly rubber and oil palm plantations, by the 1980s (Vincent and Hadi 1993; Vincent and Mohamed Ali 2005). The NFI data also distinguished five categories of virgin forests:[18] montane forests, mostly located above 1000 m, and four categories of nonmontane forests based on standing volumes of commercial timber (superior, good, moderate, poor). We aggregated the latter into two groups, better quality (superior, good) and worse quality (moderate, poor).

We estimated the forest areas for years between the NFI dates and after 2004 by linearly interpolating the natural logarithms of the NFI estimates for each catchment. The resulting estimates surely contain measurement error. Random measurement error tends to bias regression coefficients toward zero (Wooldridge 2010, pp. 78–82), and fixed effects tend to amplify this attenuation bias (Wooldridge 2010, pp. 365–368). There is no obvious reason that our interpolation procedure generated systematic measurement error that induced a bias in any particular direction. Thus, at worst, the interpolation caused our estimates of the effects of forest cover on treatment cost to be conservative: they were less likely to detect an effect if it existed, and less likely to exaggerate it if they detected it.

To construct the rainfall variable, we first divided Perak into four zones: a western coastal zone, a southeastern interior hilly zone, and two zones for the upper (northern) and lower (southern) portions of the Perak River basin, which accounts for most of the state's area (Fig. 1). The Malaysian Drainage and Irrigation Department provided data (daily in most cases, monthly in a few) for 59 rainfall stations that it operates in these zones. The data

[16] Data on other variables, such as the components of total operating cost (labor, energy, chemicals, maintenance) or quantities of specific treatment chemicals, were too incomplete to analyze.

[17] We dropped 13 observations that appeared to be affected by data-entry error, as indicated by water volumes or unit costs that were either more than twice as large or less than half as large as any other values for the same TPs. These observations came from 11 TPs. Dropping them had a negligible impact on the regression results.

[18] The NFIs contain information on the approximate time periods when logged forests were harvested. The different NFIs do not define all the time periods consistently, however. Furthermore, for time periods defined consistently between NFI 3 and NFI 4, the reported areas exhibit substantial logical inconsistencies in many catchments (e.g., the area classified as having been logged before a given year is reported as being larger in NFI 4 than in NFI 3). For these reasons, we did not analyze the effect of years since logging on treatment cost. This effect is likely negative: sediment loads decrease relatively rapidly after logging in tropical forests, though they remain higher than sediment loads from virgin forests for decades (Bruijnzeel 2004; Abdul Rahim and Zulkifli 2004).

Table 1 Descriptive statistics for the main regression sample. Unbalanced panel: 3894 monthly observations on 41 water treatment plants during 1994–2007

Variable	Mean	Min	Max	SD	
				Overall	Within
Treatment cost (RM/month, 2005 prices)[a]	34,874	1099	420,268	60,391	10,929
Water volume (m^3/month)	383,551	1792	4,492,433	713,791	153,297
Unit treatment cost (RM/m^3, 2005 prices)	0.18	0.01	2.67	0.21	0.12
Catchment area (ha)	146,848	587	929,063	315,405	0
Rainfall (mm/month)	214.5	9.4	570.3	106.4	98.8
Virgin forest (%)	40.1	0.0	95.9	28.5	4.4
Virgin forest (%): better quality	11.8	0.0	81.4	18.8	1.3
Virgin forest (%): worse quality	23.9	0.0	95.3	24.5	3.1
Virgin forest (%): montane	4.4	0.0	29.8	6.5	0.4
Logged forest (%)	36.9	2.3	100.0	22.8	4.7

[a] 2005 exchange rate: 3.79 RM/US$

ended in 2001 for the western zone and 2007 in the other zones. We calculated the simple average of the readings across the stations for a given month of a given year in a given zone, and we used these zonal rainfall values as the rainfall variables for the TPs located in the zones.

Table 1 presents descriptive statistics for the sample used in most of the regression models. This sample has somewhat fewer observations (3894) than the entire sample (4246) because it is for models that included the rainfall variables, which as just noted were not available in all years for all TPs. Comparison of the minimum values, maximum values, and overall standard deviations to the means indicates that the data show much variation. On average, forests (virgin and logged combined) accounted for a majority of the surface area of the catchments, with the area split nearly evenly between virgin and logged forests (on average, 40.1 % of catchment area vs. 36.9 %). Given our fixed-effects specification, however, the variation that matters is the variation over time within TPs, not the overall variation. Comparison of the last two columns of the table reveals that the within variation is more limited for the forest cover variables than for the other variables, especially for virgin forests and, within the virgin category, for better quality forests and montane forests. This can be expected to increase the difficulty of detecting a significant effect of those variables.

Figure 2 illustrates the magnitude of the within variation of virgin and logged forest cover across the TPs, using data from NFI 3 and NFI 4. For both forest types, forest cover decreased in some TPs and increased in others, but the general pattern was for virgin forest to decrease and logged forest to increase. A decrease in virgin area is expected in the presence of logging, as logging is what causes a virgin forest to lose its virgin status. The few TPs that show an increase in virgin area probably represent cases where some of the logged forest in their catchments was logged so long ago and had recovered so much that the NFI misclassified it as virgin. In principle, logged area can either increase or decrease, depending on the balance between additions (from the logging of virgin forests) and subtractions (due to conversion to nonforest land uses). The magnitudes of most of the changes displayed in Fig. 2 are large by developing country standards. According to UN estimates (FAO 2010, Tables 2 and 8), the percentage of the land area of Africa, South and Southeast Asia, and Central and South

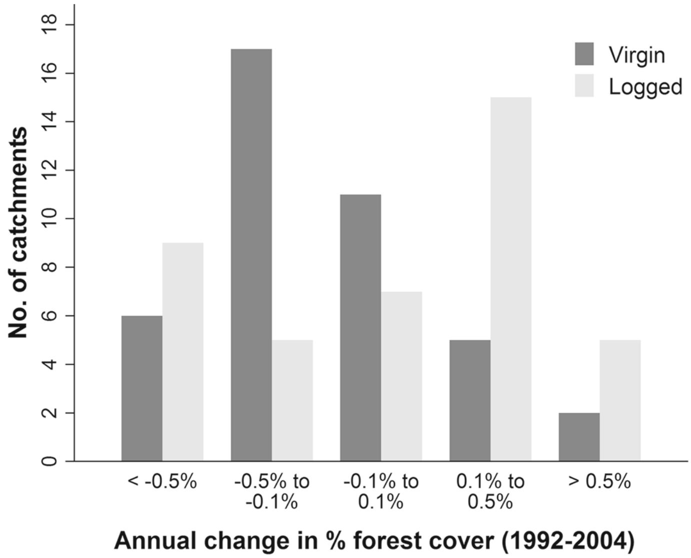

Fig. 2 Distribution of catchments for the 41 TPs in the regression sample by mean annual change in forest cover: % forest cover in 2004 minus % forest cover in 1992, divided by 12

America covered by virgin forest decreased by 0.12 %/year during 1990–2005, while the percentage covered by logged forest decreased by 0.06 %/year.[19]

Figure 3 plots the natural logarithm of unit treatment cost (constant 2005 prices; the 3rd variable in Table 1) against percent virgin forest cover (the 6th variable in Table 1) for the 3894-observation regression sample. A negative association is apparent in the figure. A similar plot for logged forest (not shown) does not exhibit this association. These relationships in the raw data suggest that virgin forests reduced treatment cost while logged forests did not, but such conclusions are premature before considering the more careful regression analysis reported next.

3 Results

3.1 Effects of Land Use on Treatment Cost

Table 2 presents our core regression results. Column (5) shows results for the fully controlled model given by Eq. (1). In this model, both virgin and logged forests have a negative effect on treatment cost, with the effect of virgin forest being larger (in absolute value) and more significant ($P = 0.013$ vs. $P = 0.027$). The difference between the coefficient estimates for the two forest types is significant at $P = 0.050$. Expressed as elasticities, the estimates indicate that avoiding the conversion of 1 % of the virgin forest in a TP's catchment to nonforest land use reduced treatment cost by 0.47 %, while avoiding conversion of 1 % of the logged forest reduced it by 0.31 %. These estimates are comparable to the elasticity of

[19] The sample for this calculation included only countries that reported a positive virgin forest area in 1990. The UN estimates refer to primary and secondary forests, which we assumed correspond to virgin and logged forests as defined in Malaysia. The endpoints, 1990 and 2005, are the UN reporting years closest to NFI 3 (1992) and NFI 4 (2004).

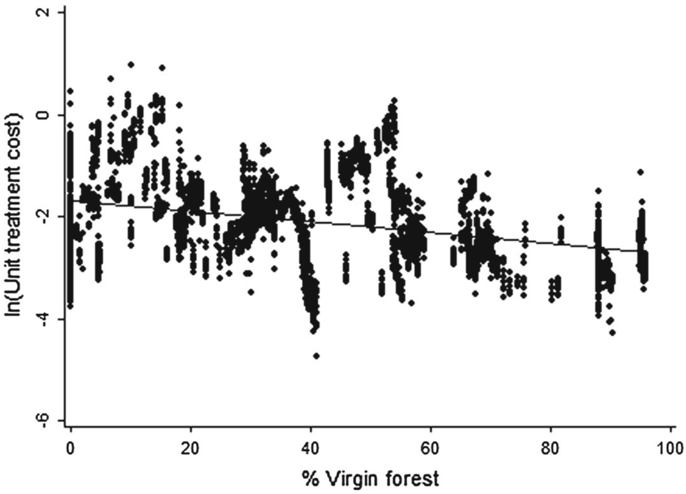

Fig. 3 Scatter plot of raw data in regression sample (3894 points). Unit treatment cost is expressed in Malaysian ringgit (2005 prices) per month. Coefficient on % Virgin forest in trend line: −0.0104 ($P = 0.007$, based on SEs clustered by TP)

treatment cost with respect to soil erosion—not forest cover—reported by Forster et al. (1987; −0.41) and the elasticities of treatment cost with respect to turbidity (a common measure of suspended solids) reported by Moore and McCarl (1987; −0.32 to −0.40), Dearmont et al. (1998; −0.27), and Murray and Forster (2001; −0.30).[20] Although the elasticities from these studies do not refer to forest cover, they are consistent with our elasticities in suggesting that environmental quality, however defined, has a less-than-proportional effect on treatment cost.

Converting our elasticities to semielasticities facilitates comparison to other published results. Using notation from Sect. 2.1, avoiding the conversion of 1 % of a TP's catchment from forest to nonforest land use changes expected treatment cost by $\hat{\beta}/(1 + F/A)$ %. At mean values in our sample, avoiding the conversion of 1 % of a catchment from forest to nonforest land use reduced treatment cost by 1.16 % in the case of virgin forest and 0.85 % in the case of logged forest. These estimates are much smaller than the semielasticity reported by Ernst (2004) for US TPs, a 2 % reduction in cost for an additional 1 % of a TP's catchment being forested. This discrepancy could be due to differences between tropical Malaysia and the temperate US or, more likely in our view, the lack of control for confounding factors in Ernst's analysis.

Results for models (1)–(4) reveal that a negative effect of forest cover, especially virgin forest cover, is evident even in models that are not fully controlled. Model (1) is the least controlled model. It includes just an aggregate forest cover variable (the percent of a TP's catchment in virgin and logged forest combined) and water volume, without rainfall or fixed effects of any kind. This pooled OLS model estimates the effect of forest cover using all the variation in the data, both between TPs and within TPs. The estimated effect is negative but not very significant. Models (2)–(4) allow the effects of virgin forest and logged forest to differ, and they progressively add controls for potentially confounding factors, first rainfall

20 Holmes (1988) reported a much lower turbidity elasticity, −0.07.

Table 2 Robustness of estimated effects of forest cover on water treatment cost

Variables[a]	Aggregated forest: no controls	Disaggregated forest: no controls	Add rainfall	Add FEs[b] for WTPs	Add FEs for time	IV[c]	Annual
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln (1 + Forest share)	−1.13						
	(0.134)						
ln (1 + Virgin forest)		−1.02	−1.01*	−1.35***	−1.63**	−1.63**	−1.57**
		(0.101)	(0.097)	(0.001)	(0.013)	(0.011)	(0.024)
ln (1 + Logged forest)		−0.173	−0.247	−0.974**	−1.17**	−1.21**	−1.07*
		(0.807)	(0.720)	(0.011)	(0.027)	(0.022)	(0.073)
ln (Water volume)	0.648***	0.653***	0.656***	0.146**	0.154**	0.173*	0.162**
	(0.000)	(0.000)	(0.000)	(0.030)	(0.028)	(0.051)	(0.042)
ln (Rainfall)			0.290	−0.259**	−0.256**	−0.251**	−1.27
			(0.281)	(0.037)	(0.014)	(0.018)	(0.462)
ln (Rainfall), squared			−0.0449	0.0255**	0.0239**	0.0234**	0.103
			(0.133)	(0.032)	(0.016)	(0.020)	(0.519)
FEs: TPs	No	No	No	Yes	Yes	Yes	Yes
FEs: TP upgrading	No	No	No	Yes	Yes	Yes	Yes
FEs: Years	No	No	No	No	Yes	Yes	Yes
FE: Months	No	No	No	No	Yes	Yes	No
Observations	3894	3894	3894	3894	3894	3723	364
R^2	0.674	0.691	0.696	0.954	0.956	0.956	0.977

Time span: 1994–2007

$^{***}P < 0.01$; $^{**}P < 0.05$; $^{*}P < 0.1$

[a] P values are shown in parentheses below the coefficients. They refer to two-sided t tests of the null hypothesis that a coefficient equals zero, and they are based on robust standard errors clustered by TP (41 clusters)

[b] *FE* fixed effects

[c] ln(Water quantity) was instrumented using its 2-month lag

and then the fixed effects.[21] The effect of virgin forest remains negative and larger and more significant than the effect of logged forest as these controls are added.

Model (6) is an IV version of the fully controlled model that treats water volume as endogenous,[22] with the second lag of water volume used as an instrument. This variable satisfies the exclusion restriction—there is no apparent reason for it to influence current treatment cost—and it is an extremely strong instrument ($F = 1399$ in the first-stage regression). We chose the second lag instead of the first lag to guard against correlation with the error term, if the latter is serially correlated.[23] As can be seen, the estimated effects of virgin and logged forests differ little from those in model (5). Moreover, Wooldridge's (1995) robust score test indicates that the IV correction is unnecessary, as the test does not reject the null that water volume is exogenous ($P = 0.464$).[24]

Models (1)–(6) address the Moulton problem by clustering the standard errors by TP. Clustering is an asymptotic correction that requires a large number of clusters, with 40–50 clusters being the rule-of-thumb (Angrist and Pischke 2009, Ch. 8). The number of clusters in the models was 41. The importance of clustering can be revealed by comparing the TP-clustered standard errors for the virgin and logged forest variables in model (5) to alternative standard errors for these variables (Table 3). Consider the virgin forest variable. The TP-clustered standard error for this variable is 0.628 (the first row), which provided the basis for the P value reported in Table 2. The conventional robust standard error is less than half as large (the second row). Using the latter would exaggerate the significance of the effect of virgin forest on treatment cost. As discussed in Sect. 2.1, TP-clustered standard errors address both the Moulton problem and serial correlation. The Moulton problem is the more important source of the discrepancy between the clustered and conventional robust standard errors: the Newey-West standard error (the third row), which addresses heteroskedasticity and serial correlation but not the Moulton problem, is not much larger than the conventional robust standard error, whereas the standard error clustered by TP-year instead of TP (the fourth row), which addresses the Moulton problem and serial correlation within years but not between years, is not much smaller than the TP-clustered standard error. The same patterns held for the standard errors for the logged forest variable. Addressing the Moulton problem is obviously necessary.

Model (7), which like models (5) and (6) is a fully controlled model, addresses the Moulton problem in a different way: it eliminates the repeated observations on land use across months within a given year by collapsing all the data to annual. The effects of the forest variables

[21] Fixed effects were more appropriate than random effects because the sample included nearly all TPs in Perak, not a random selection of them (Wooldridge 2010, pp. 285–287; Kennedy 2008, p. 291). Moreover, Hausman tests rejected the null that the random effects were uncorrelated with the error term: $\chi^2(24) = 95.3$ ($P = 0.000$) if the test used the disturbance variance estimate from the random effects model and $\chi^2(24) = 97.0$ ($P = 0.000$) if it used the estimate from the fixed effects model. Despite these test results, the random effects coefficient estimates were not very different from the fixed effects estimates given by model (5) in Table 2: -1.41 ($P = 0.000$) for virgin forest and -1.06 ($P = 0.000$) for logged forest.

[22] Endogeneity of a different sort—a correlation between the forest cover variables and the error term—could occur if the four TPs that were established after 1994 were designed to cope with the effects of the land-use changes that occurred during the sample period. Reestimating model (5) with those TPs excluded from the sample did not change the results much: the coefficients on virgin forest and logged forest were -1.56 ($P = 0.019$) and -1.13 ($P = 0.0.34$), respectively.

[23] Results are very similar to those for model (6) if we use the first lag as an IV: the coefficients on virgin forest and logged forest are -1.62 ($P = 0.009$) and -1.18 ($P = 0.020$), respectively. The instrument is again very strong ($F = 1989$).

[24] This is also the case if we use the first lag as an instrument ($P = 0.711$).

Table 3 Robustness checks for model (5) in Table 2: alternative standard errors (SEs)

Model description	Coefficient estimates and SEs for	
	ln(1 + Virgin share)	ln(1 + Logged forest)
Model (5) from Table 2	−1.63**	−1.17**
	(0.628)	(0.511)
Robust SEs, no clustering	−1.63***	−1.17***
	(0.258)	(0.229)
Newey-West SEs	−1.63***	−1.17***
	(0.304)	(0.265)
Robust SEs, clustered by TP-year	−1.63***	−1.17**
	(0.540)	(0.491)
Robust SEs, clustered by zone-year	−1.63**	−1.17**
	(0.702)	(0.526)
PCSE[a] without TP correlations	−1.66***	−1.14***
	(0.405)	(0.382)
PCSE with TP correlations	−1.66***	−1.14***
	(0.441)	(0.402)
Model (5) from Table 2, excluding 6 TPs with large catchments	−1.47**	−1.07**
	(0.625)	(0.491)

*** $P < 0.01$; ** $P < 0.05$; * $P < 0.1$
[a] Panel-corrected SEs (Beck and Katz 1995)

remain negative in this model and are only slightly smaller than the effects in model (5). Their significance is somewhat lower, especially for the logged forest variable.

The Moulton problem results from the error term not being independently distributed across months in a given year for a given TP. Another potential source of nonindependence is correlation of errors between TPs (i.e., spatial correlation). We investigated this in three ways, and all indicated that correlations between TPs were not large. First, we clustered the standard errors by rainfall zone-year (the fifth row in Table 3), to allow correlations between TPs in the same zone. The estimates were only slightly larger than the TP-clustered estimates. Second, we estimated the fully controlled model using Beck and Katz's (1995) panel-corrected standard errors, which can correct for correlations between TPs in addition to heteroskedasticity and serial correlation. The panel-corrected standard errors were only slightly larger with the correction for TP correlations (the seventh row) than without it (the sixth row). Third, we estimated the fully controlled model for a sample that excluded six TPs on lower to middle stretches of the Perak River (Fig. 1).[25] The catchments for these TPs include the catchments for several other TPs nested within them, which could cause correlated errors. Omitting the six TPs reduced the number of observations to 3242, but it barely changed the TP-clustered standard errors for either forest variable. It did not change the coefficients for these variables much either.

Our data enabled us to estimate two variations on Eq. (1) that included more complex specifications of the effect of forest cover. The results did not indicate any advantage to using these models instead of model (5) and thus are not shown. The first variation tested for differences across types of virgin forest by including variables for better quality and montane

[25] Air Ganda, BB Seri Iskandar, Kg. Gajah, Kg. Paloh, Kota Lama Kiri, and Teluk Kepayang.

virgin forests in addition to the overall virgin forest variable. Among the three virgin forest types, montane forests tend to be found on the steepest slopes and better quality forests on the least steep slopes. This topographical pattern implies a negative coefficient on montane forests (protection against a higher risk of soil erosion) and a positive coefficient on better quality forests (protection against a lower risk). The coefficients on the two additional virgin forest variables were not significant ($P > 0.6$), however, which suggests that their effects do not differ from the effect of the reference, worse quality category. As Table 1 shows, however, these two types vary less over time than does the worse quality type. We cannot rule out the possibility that their insignificance resulted from this limited within variation.

The second variation allowed the effect of forest cover to vary with catchment size, by interacting the forest cover variables with catchment area. The rationale for this interaction is evidence from the forest hydrology literature that sediment yield per unit area decreases with catchment area, as larger catchments trap a larger share of sediment upstream in river channels and valleys (Dunne and Leopold 1978; Bruijnzeel 2004; Abdul Rahim and Zulkifli 2004).[26] This redeposition effect implies a positive coefficient on the interaction terms. That is what we found, but neither of the interaction terms was very significant ($P > 0.2$).

3.2 Effects of Water Volume and Rainfall on Treatment Cost

Although our focus is on the effects of forest cover, results for the other variables offer evidence on the credibility of our models. Prior research suggests that water volume should have a positive, less-than-proportional effect on cost, and that is what we find in model (5) in Table 2. With cost and water volume both being log-transformed, the coefficient on water volume is directly interpretable as an elasticity, and its value is 0.154 ($P = 0.028$). This is much smaller than the elasticities reported by econometric studies on water treatment cost in the US, which range from 0.66 to 0.96 (Forster et al. 1987; Holmes 1988; Dearmont et al. 1998; Murray and Forster 2001; Piper 2003). An explanation for this discrepancy is suggested by results for models (1)–(3) in Table 2, which have elasticities closer to those in the US studies, 0.648–0.656. These models exclude TP fixed effects; hence, their elasticities are influenced by variation between TPs, not just within TPs as in model (5). The samples for the US studies were all either cross-sections or short panels, and none of the panel studies included TP effects; hence, the US elasticities are also influenced by variation between TPs. The explanation is thus that returns to scale in water treatment are stronger within individual TPs, which is implied by the lower elasticity in model (5), than between TPs, which is implied by the larger elasticities in models (1)–(3) and the US studies.[27]

Results for the rainfall variables in model (5) are consistent with our hypothesis of a nonlinear effect that is negative at lower levels of rainfall (cost is decreasing in rainfall, due to a stronger dilution effect) and positive at higher levels (cost is increasing, due to a stronger soil erosion effect).[28] The turning point is at 212 mm/month, which is the 56th percentile

[26] Table 1 in FAO and CIFOR (2005) indicates that land use affects sediment loads only in catchments up to 10,000 ha, which is smaller than 40 % of the observations in our sample. Ogden et al. (2013) note that the table is drawn from a source that provides no citations to support this threshold.

[27] An additional explanation could be that the US elasticities are biased upwards by a correlation between water volume and some omitted variable. A strong candidate for such a variable is catchment area. In our sample, water volume has a relatively large positive correlation with catchment area ($\rho = 0.310,\ P = 0.000$). This correlation is not surprising; larger TPs require larger supplies of source water and thus have larger catchments. The US studies do not control for catchment area, but model (5) controls for it through the TP effects.

[28] The sign pattern on the linear and quadratic terms is the opposite (positive and negative) if TP effects are excluded, as in model (3). This demonstrates the importance of controlling for unobserved TP characteristics.

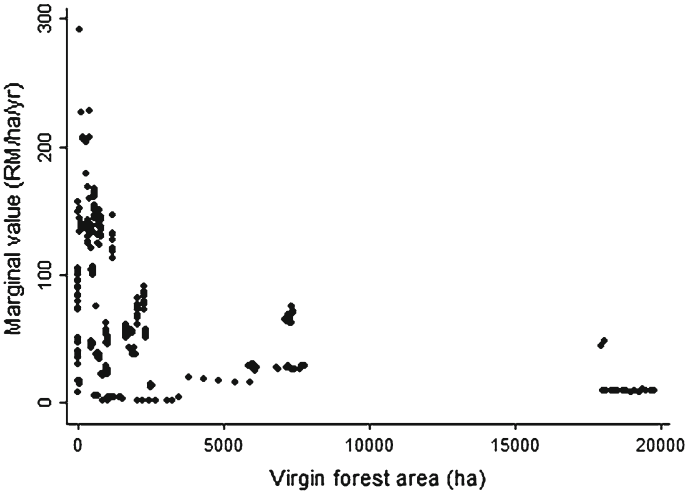

Fig. 4 Marginal value of water purification service supplied by virgin forests to TPs in Perak during 1994–2007. Each point refers to a particular TP in a particular year (304 points)

of the rainfall distribution in the sample. The sign pattern is retained but the effects are less significant in the annual model in column (7), which implies that intraannual variation in rainfall affects cost much more than interannual variation does.

3.3 Marginal Value of the Water Purification Service

Figure 4 displays the marginal value of the water purification service supplied by virgin forests to TPs in Perak during 1994–2007. The marginal value is defined as the annual reduction in treatment cost that results from avoiding the conversion of 1 ha of virgin forest to a nonforest land use. Monthly estimates were calculated by applying Eq. (3) to data in the regression sample for model (5) in Table 2 and then converted to annual estimates by multiplying the mean across the months in a given TP-year by 12. The figure does not show the marginal values for six TPs with very large catchments ($>$ 400,000 ha; the next largest in the sample is 51,000 ha), which are off the scale to the right. The mean marginal value for them, RM1.81/ha/year, is much smaller than the mean marginal value for the 35 TPs shown in the figure, RM67.75/ha/year.

Two features of the figure are notable. One is that the scatter trends downward, which is the expected shape for a textbook marginal benefit curve. This shape results from the derivative of Eq. (3) with respect to virgin forest area—i.e., $\partial^2 \mathrm{E}\left(C \,\middle|\, \mathbf{L}, Q, R.c_i, \theta_y, \theta_m\right)/\partial F^2$—being positive. It implies that cost-minimizing TPs would pay diminishing amounts to protect additional hectares of virgin forest: their demand curve for the water purification service is downward-sloping. The second feature is more striking, and it is that the marginal values vary greatly for a given area of virgin forest. This mainly reflects variation between TPs: the water purification service is much more valuable to some TPs than to others. Variation in marginal values over time for a given TP is a secondary explanation.

The magnitude of these values can be gauged in several ways. One is to consider the implications for the cost of operating a TP if it had to pay for the water purification service, just as it pays for labor, energy, chemicals, and maintenance. This cost can be estimated by multiplying the marginal values in Fig. 4 by the corresponding virgin forest areas, and then expressed as a ratio by dividing by the actual operating cost for each TP in each year. The mean of this ratio was 0.417 (SD = 0.273): on average, the value of water purification provided by virgin forests, valued on the margin, was equivalent to more than a third of TPs' expenditures on priced inputs. The maximum value of the ratio in the sample was 1.20; that particular TP received a water purification service whose value was greater than its expenditure on priced inputs. Clearly, virgin forests provided a valuable service to the TPs.

A second way to gauge the magnitude of the service values is to compare them to the opportunity cost of protecting virgin forests from conversion. Malaysian input–output tables report aggregate operating surplus for rubber and oil palm plantations in 2005 (Department of Statistics 2010). This is a good measure of short-run profits in those sectors, albeit for all of Malaysia and not specifically for Perak. The input–output tables also report taxes paid by the sectors, which can be added to operating surplus to form a more complete measure of producer surplus in the sectors. These values can then be divided by aggregate plantation areas in 2005 (provided by the Department of Statistics) to calculate producer surplus per ha. The resulting values are RM2,360/ha/year for rubber and RM2812/ha/year for oil palm. Each of these is more than 30 times the mean value of the water purification service shown in Fig. 4. Taken at face value, this suggests that the value of the water purification service is insufficient to justify protecting virgin forests against conversion to plantation crops.

The value of the water purification service can also be compared to the opportunity cost of protecting virgin forests against logging instead of conversion. The relevant measure of value in this case is the difference in service values between virgin and logged forests, which is given by

$$-\frac{\left(\hat{\beta}^V - \hat{\beta}^L\right) \cdot \mathrm{E}\left(C \left| \mathbf{L}, Q, R, c_i, \theta_y, \theta_m \right.\right)/A}{1 + F^V/A}, \tag{4}$$

where V and L distinguish the coefficients on virgin and logged forests, respectively. For the TPs and years in Fig. 4, this expression has a mean of RM19.06/ha/year (SD = RM14.78/ha/year; range RM0.41– 81.87/ha/year). All forests in Malaysia are owned by state governments, and the loss of fiscal revenue from royalties and other timber fees is a major opportunity cost of protection against logging that has concerned the Perak state government (Schwabe et al. 2014). Data from the Forestry Department Peninsular Malaysia indicate that this revenue was RM2831.88/ha logged in the state in 2005. The Department recommends a 30-year harvest cycle, which implies an annual timber revenue flow of RM94.40/ha/year if the state's forests were managed sustainably in 2005. Although this is greater than the mean difference in service values, the gap is much smaller than in the case of protecting virgin forests against conversion to plantation crops.

4 Discussion and Conclusions

We have presented significant, robust evidence that protecting virgin and logged forests against conversion to nonforest land uses can reduce the operating costs of water treatment plants. This evidence is based on Malaysian panel data, which enabled us to use fixed-effects regression models to control more completely for potentially confounding factors than previous studies have done. Our estimates indicate that forests have an inelastic effect on treatment cost, with virgin forests having a larger effect than logged forests. We detected these effects despite using interpolated estimates of forest cover, which likely biased the estimated effects toward zero, and despite the relatively limited amount of variation in forest cover over time, especially for virgin forests.

Valuing virgin forests at the marginal reductions in treatment costs that they generated indicates that the forests' value to treatment plants was equivalent, on average, to more than a third of the plants' aggregate expenditures on priced inputs (labor, energy, chemicals, maintenance). Our estimate of the marginal value of this service is very small, however, compared to producer surpluses for rubber and oil palm plantations in 2005, which we used to estimate the opportunity cost of protecting virgin forests against conversion. For multiple reasons, we believe that this comparison exaggerates the gap between the service's marginal benefit and its marginal cost. Regarding the marginal benefit, as just mentioned our estimate of the effect of virgin forest on treatment cost was probably biased downward due to measurement error, and as discussed earlier our analysis did not account for reductions in capital costs or the net value of additional treated water that treatment plants might have produced as a result of having cleaner source water.

Regarding the marginal cost of the service, our producer surplus estimates for rubber and oil palm are short-run estimates, which represent an amalgam of returns to all fixed factors of production, not only land. Establishing rubber and oil palm plantations requires substantial investments in infrastructure, terracing, and drainage. A proper marginal cost estimate for the water purification service would include only land rent, which is net of returns on these other investments. Moreover, our producer surplus estimates are averages across all existing plantations in Malaysia, not estimates for marginal plantations in Perak. Most plantations in Malaysia are in the lowlands, which offer superior conditions for growing rubber and oil palm. These plantations can be expected to generate larger surpluses than new plantations in the hilly and mountainous regions of Perak where most virgin forests remain. For these two reasons, our producer surplus estimates overstate the marginal cost of the service by amounts that are unknown but surely large.

For similar reasons, we have underestimated the marginal benefit and overestimated the marginal cost of the enhanced water purification service generated by protecting virgin forests against logging. Our estimate of the average economic shortfall for this type of protection was already much smaller than the average shortfall for protecting virgin forests against conversion to rubber or oil palm. This suggests that water purification services are more likely to justify protection of virgin forests in locations where the main threat comes from logging (i.e., degradation), not conversion to plantation crops (deforestation).

We close by emphasizing two points. The first is that averages across treatment plants do not tell the whole story, or even the most important part of the story. Figure 4 demonstrates that the marginal value of the water purification service exhibits great heterogeneity; hence, the marginal benefits of virgin-forest protection are much greater in some locations than in others. If the marginal costs of protection also exhibit heterogeneity, as they surely do, then it is possible that protection could be justified even in locations where marginal benefits are low (if marginal costs are even lower) and not justified in locations where marginal benefits are high (if marginal costs are even higher).

The second point is that water purification is just one of many watershed services and other ecosystem services that tropical forests can supply (Pattanayak 2004; Pattanayak and Wendland 2007). In Malaysia, recent research indicates that tropical forests can also mitigate floods (Tan-Soo et al. 2014) and generate large values related to biodiversity conservation (Vincent et al. 2014). Protection decisions that appear to be economically unjustified when only water purification is considered might turn out to be justified when a broader range of services is taken into account.

Acknowledgments This study was funded by the Global Environment Facility through the United Nations Development Programme (MAL/04/G31), with additional support from the Government of Malaysia (through the Ministry of Natural Resources and Environment and the Forest Research Institute Malaysia) and the Center for International Forestry Research (CIFOR). The cooperation of numerous Malaysian government agencies is gratefully acknowledged: the Forestry Department Peninsular Malaysia; the Departments of Agriculture, Drainage and Irrigation, and Statistics; the National Water Services Commission; and the Perak Water Board. We also thank David James and two anonymous reviewers for helpful comments.

References

Abdul Rahim AS, Mohd Shahwahid HO (2011) A panel data analysis of timber harvesting operations and its impact on the cost of water treatment. Aust J Basic Appl Sci 5:598–601

Abdul Rahim N, Zulkifli Y (2004) Hydrological impacts of forestry and land-use activities: Malaysian regional experience. In: Abdul Rahim N (ed) Water: forestry and land use perspectives, Technical Documents in Hydrology No. 70, International Hydrological Programme. UNESCO, Paris, pp 86–105

Alcott E et al (eds) (2013) Natural and engineered solutions for drinking water supplies: lessons from the northeastern United States and directions for global watershed management. CRC Press, Boca Raton

Angrist JD, Pischke JS (2009) Mostly harmless econometrics: an empiricist's companion. Princeton University Press, Princeton

Barten PK, Ernst CE (2004) Land conservation and watershed management for source protection. J Am Water Works Assoc 96:121–135

Beck NL, Katz JN (1995) What to do (and not to do) with time-series cross-section data. Am Polit Sci Rev 89:634–647

Bennett G et al (2013) Charting new waters. Forest Trends, Washington

Bruijnzeel LA (2004) Hydrological functions of tropical forests. Agric Ecosyst Environ 104:185–228

Carlson KM et al (2014) Influence of watershed-climate interactions on stream temperature, sediment yield, and metabolism along a land use intensity gradient in Indonesian Borneo. J Geophys Res Biogeosci 119:1110–1128

Chichilnisky G, Heal GM (1998) Economic returns from the biosphere. Nature 391:629–630

Conte M et al (2011) Retention of nutrients and sediment by vegetation. In: Kareiva P et al (eds) Natural capital: theory and practice of mapping ecosystem services. Oxford University Press, Oxford, pp 89–110

Dearmont D et al (1998) Costs of water treatment due to diminished water quality: a case study in Texas. Water Resour Res 34:849–853

Department of Statistics (2010) Input-output tables, Malaysia 2005: volume 2. Putrajaya, Malaysia

Dissmeyer GE (ed) (2000) Drinking water from forests and grasslands: a synthesis of the scientific literature. Gen Tech Rep SRS-39, USDA Forest Service, Southern Research Station, Asheville, North Carolina

Dudley N, Stolton S (2003) Running pure: the importance of forest protected areas to drinking water. World Bank and WWF, Washington

Dunne T (1979) Sediment yield and land use in tropical catchments. J Hydrol 42:281–300

Dunne T, Leopold LB (1978) Water in environmental planning. WH Freeman, New York

Ernst C (2004) Protecting the source: land conservation and the future of America's drinking water. Trust for Public Land, San Francisco

Ernst C et al (2004) Protecting the source: conserving forests to protect water. Opflow 30(1):4–7

FAO (2010) Global forest resources assessment 2010: main report. Food and Agriculture Organization of the United Nations, Rome

FAO and CIFOR (2005) Forests and floods. Food and Agriculture Organization of the United Nations, Rome and Center for International Forestry Research Bogor, Indonesia

Ferraro PJ et al (2012) Forest figures: a review of ecosystem services valuation and policies in developing countries. Rev Environ Econ Policy 6:20–44

Forster DL et al (1987) Soil erosion and water treatment costs. J Soil Water Conserv 42:349–352

Freeman AM III et al (2014) The measurement of environmental and resource values: theory and methods, 3rd edn. RFF Press, Taylor & Francis, Oxford

Freeman J et al (2008) Statistical analysis of drinking water treatment plant costs, source water quality, and land cover characteristics. Trust for Public Land, San Francisco

Gartner T et al (eds) (2013) Natural infrastructure: investing in forested landscapes for source water protection in the United States. World Resources Institute, Washington

Hewlett JD (1982) Principles of forest hydrology. University of Georgia Press, Athens

Holmes TP (1988) The offsite impact of soil erosion on the water treatment industry. Land Econ 64:356–366

Huber PJ (1967) The behavior of maximum likelihood estimates under nonstandard conditions. In: Neyman LM (ed) Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol 1. University of California Press, Berkeley, pp 221–233

Hurley T, Mazumder A (2013) Spatial scale of land-use impacts on riverine drinking source water quality. Water Resour Res 49:1591–1601

Kennedy P (2008) A guide to econometrics, 3rd edn. Blackwell, Malden

Kenny A (2006) Ecosystem services in the New York City watershed. http://www.ecosystemmarketplace.com/pages/dynamic/article.page.php?page_id=4130&section=home&eod=1_

Krishnaswamy J et al (2001) Spatial patterns of suspended sediment yields in a humid tropical watershed in Costa Rica. Hydrol Process 15:2237–2257

Lembaga Air Perak (n.d.) Laporan tahunan 2010. Ipoh, Perak, Malaysia

McConnell KE, Bockstael NE (2005) Valuing the environment as a factor of production. In: Mäler KG, Vincent JR (eds) Handbook of environmental economics, vol 2. North-Holland, Amsterdam, pp 621–669

McDonald R, Shemie D (2014) Urban water blueprint: mapping conservation solutions to the global water challenge. The Nature Conservancy, Washington

Mohd Akbar HJ, Rusnah R (2004) Water resources management for domestic and industrial needs. In: Abdul Rahim N (ed) Water: forestry and land use perspectives, Technical Documents in Hydrology No. 70, International Hydrological Programme. UNESCO, Paris, pp 1–15

Mohd Nordin A et al (2000) The effects of raw water turbidity in water treatment process at Sg. Kampar treatment plant. In: Proceedings of environmental seminar, School of Chemical Engineering. Universiti Sains Malaysia, Penang, pp 46–51

Moore WB, McCarl BA (1987) Off-site costs of soil erosion: a case study in the Willamette Valley. West J Agric Econ 12:42–49

Moulton BR (1986) Random group effects and the precision of regression estimates. J Econom 32:385–397

Murray C, Forster L (2001) A study of pesticide use, farming practices, and community drinking water treatment costs in the Maumee and Great Lakes Basins. Rep. Series AEDE-RP-0013-01, Agricultural, Environmental, and Development Economics, Ohio State University, Columbia, Ohio

Ogden FL et al (2013) Effect of land cover and use on dry season river runoff runoff efficiency and peak storm runoff in the seasonal tropics of Central Panama. Water Resour Res 49: doi:10.1002/2013WR013956

Pattanayak SK (2004) Valuing watershed services: concepts and empirics from southeast Asia. Agric Ecosyst Environ 104:171–184

Pattanayak SK, Wendland KJ (2007) Nature's care: diarrhea, watershed protection and biodiversity conservation in Flores, Indonesia. Biodivers Conserv 16:2801–2819

Pattanayak SK, Kramer RA (2001) Worth of watersheds. Environ Dev Econ 6:123–146

Piper S (2003) Impact of water quality on municipal water price and residential water demand and implications for water supply benefits. Water Resour Res 39: doi:10.1029/2002WR001592

Postel SL, Thompson BH Jr (2005) Watershed protection: capturing the benefits of nature's water supply services. Nat Resour Forum 29:98–108

Sagoff M (2002) On the value of natural ecosystems. Polit Life Sci 21:19–25

Schwabe KA et al (2014) Creation of Malaysia's Royal Belum State Park: a case study of conservation in a developing country. J Environ Dev. doi:10.1177/1070496514551173

Sklenar K et al (2012) Source water protection vision and roadmap. PDF Report #4176a, Water Research Foundation, Denver, Colorado

Spiller M et al (2013) Pollution source control by water utilities: characterisation and implications for water management. Water Environ J 27:177–186

Sthiannopkao S et al (2007) Soil erosion and its impacts on water treatment in the northeastern provinces of Thailand. Environ Int 33:706–711

Tan-Soo JS et al (2014) Econometric evidence on forest ecosystem services: deforestation and flooding in Malaysia. Environ Resour Econ 59. doi:10.1007/s10640-014-9834-4

Vincent JR (2011) Valuing the environment as a production input. In: Haque AKE et al (eds) Environmental valuation in South Asia. Cambridge University Press, New Delhi, pp 36–77

Vincent JR, Hadi Y (1993) Malaysia. In: Council National Research (ed) Sustainable agriculture and the environment in the humid tropics. National Academy Press, Washington

Vincent JR, Mohamed Ali R (2005) Managing natural wealth: environment and development in Malaysia. Resources for the Future, Washington

Vincent JR et al (2014) Tropical countries may be willing to pay more to protect their forests. Proc Natl Acad Sci USA 111:10113–10118

White HL (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48:817–838

Whittington D (2010) What have we learned from 20 years of stated preference research in less-developed countries? Annu Rev Resour Econ 2:209–236
Wooldridge JM (1995) Score diagnostics for linear models estimated by two stage least squares. In: Maddala GS et al (eds) Advances in econometrics and quantitative economics: essays in honor of Professor CR, Rao edn. Blackwell, Oxford, pp 66–87
Wooldridge JM (2010) Econometric analysis of cross section and panel data, 2nd edn. MIT Press, Cambridge
Zeger SL, Liang KY (1986) Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42:121–130