MAGKS

Joint Discussion Paper
Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 18-2019

Victor von Loessl and Heike Wetzel

Revenue decoupling and energy consumption: Empirical evidence from the U.S. electric utilities sector

Coordination: Bernd Hayo • Philipps-University Marburg
School of Business and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Revenue decoupling and energy consumption: Empirical evidence from the U.S. electric utilities sector

Victor von Loessl[a,*], Heike Wetzel[a]

[a] *Institute of Economics, University of Kassel, Nora-Platiel-Str. 4, 34109 Kassel, Germany*

Abstract

Energy efficiency provides a substantial opportunity to tackle increasing greenhouse gas emissions. However, in traditionally regulated energy markets, energy providers maximize their profits by selling electricity or heat as long as their marginal revenue exceeds their marginal costs of production. This so called 'throughput incentive' fundamentally restricts the motivation of utilities to invest in energy efficiency. This paper therefore investigates the relation between the regulatory policy revenue decoupling, that separates utilities' revenue from sales fluctuations, and electricity customers' energy demand and efficiency in the U.S. To address the research question at hand, we follow recent developments in energy demand function modeling and Stochastic Frontier Analysis (SFA) estimation techniques that allow to account for persistent as well as transient efficiency. The estimation results show a significant negative correlation between revenue decoupling and electricity consumption patterns. Furthermore, we find electricity customers have small transient inefficiency. However, results indicate an underlying persistent inefficiency across the entire electric sector.

Keywords: Revenue decoupling, energy efficiency, stochastic frontier analysis, demand frontier function, transient and persistent efficiency

JEL classification: C23, L51, L94

*Corresponding author:
vonloessl@uni-kassel.de, +49 561 804 7706 (Victor von Loessl)

1. Introduction

Rising global energy demand is one well-documented driver of greenhouse gas emissions (see e.g. Ang (2007) or Apergis and Payne (2010)). Thus, the ongoing global increase in energy demand (OECD, 2012) lies at odds with climate protection goals, demonstrating a need for action in the energy sector. Even though CO_2 emissions in the U.S. were about 10.4 % lower in 2015 than in the peak year 2007 (EPA, 2017) the U.S., as the 2nd-largest CO_2 emitter in the world, still substantially contribute to global CO_2 emissions. Furthermore, the set target of the Obama administration to reduce emissions by 26 - 28 % by 2025 compared to 2005 levels seems ambitious and potentially unachievable (Victor et al., 2017). While several aspects need to be taken into account, the U.S. Environmental Protection Agency (EPA) as well as the International Energy Agency (IEA) highlight the importance of energy efficiency as an energy resource and as a key factor in reducing CO_2 emissions (EPA, 2008; IEA, 2009).

To support the efficiency of climate protection measures, Victor et al. (2017) identify the need for more adequate indicators that evaluate the actual impact of regulatory policies. We take up their call and investigate the relation between an energy market regulation, namely revenue decoupling, and energy demand and efficiency.

In traditionally regulated electricity markets electricity providers maximize their profit by selling as much electricity as possible, as long as the marginal costs of production are below the marginal revenue (Eto et al., 1997). Considering this so called 'throughput incentive', electricity utilities have little motivation to invest in energy efficiency, as a reduction of sold electricity implies a reduction of revenue (this also applies for the gas sector) (Kahn-Lang, 2016). Alongside other policies the concept of decoupling was implemented for investor-owned energy utilities to tackle this misleading and unsustainable incentive of sales-maximization (NARUC, 2007). In a nutshell, decoupling is the separation of the total electricity delivered by an electric utility and its profits (Brennan, 2010), thereby making revenue independent from sales fluctuations (NARUC, 2007). By 2015, decoupling was implemented in 26 U.S. states and the District of Columbia. Thirteen of those states and the District of Columbia have implemented decoupling policies in the electricity sector (EEI, 2015). Recognizing revenue decoupling as a major policy in energy market regulations, the question arises whether it has a significant impact. As this study only examines the electricity sector, the aim of this paper is to investigate the relation between revenue decoupling and the electricity demand and efficiency of electricity utilities' customers in the U.S.

Voigt et al. (2014) argue that energy intensity, generally defined as the ratio of energy consumption to gross domestic product (GDP), would be the reciprocal of energy efficiency and thus a valid measurement. However, Filippini and Hunt (2015*a*) argue that energy intensity is solely the reciprocal of energy productivity but not an optimal measurement for energy efficiency[1]. They compare average energy intensity scores with persistent energy efficiency levels via an estimated energy demand function for 49 U.S. states and show that energy intensity is only a valid proxy for energy efficiency in some cases and should therefore be used carefully.

One common approach is to decompose energy intensity into several components, such as changes in fuel sources, alterations of the production level, changes of the overall structure of the economy and energy efficiency (Jimenez and Mercado, 2014). Comprehensive methodological reviews about decomposition approaches are provided by Ang et al. (2010) and Boyd and Roop (2004).

Alternatively, different frontier analysis approaches have been conducted to estimate energy efficiency levels (Filippini and Hunt, 2012). Usually frontier analysis approaches are either non-parametric, such as Data Envelopment Analysis (DEA), or parametric, such as Stochastic Frontier Analysis (SFA). As specified in detail in Section 3, this paper will follow the argumentation of Filippini and Hunt (2015*a*) to understand energy (in)efficiency based on the micro-economic theory of production. We pursue their suggestion to use a SFA approach as originally introduced by Aigner et al. (1977). Furthermore, the differentiation between persistent and transient efficiency (long and short-term) recently gained attention in the general debate on energy efficiency (Filippini and Hunt, 2015*a*) and will thus play an important role in the estimation approach of this paper.

The basis for our analysis is an unbalanced panel data set for the period of 2001 to 2015, which includes utility specific information regarding total sales, absolute customer numbers and consumption shares (commercial, industry and households) as well as state-level data regarding Cooling Degree Days (CDD), Heating Degree Days (HDD), real Gross Domestic Product (GDP) and average electricity sales prices. Furthermore, we control whether a utility is affected by Energy Efficiency Resource Standards (EERS) which are implemented on the state level and, of course, whether a utility implemented a decoupling-mechanism or not. Throughout all model specifications, our analysis reveals a statistically highly significant negative correlation between revenue decoupling true up plans and electricity consumption.

[1]The U.S. Department of Energy discusses this matter at: https://energy.gov/eere/analysis/energy-intensity-indicators-efficiency-vs-intensity, stating that energy intensity can be a qualitative proxy for energy efficiency if other explanatory variables are accounted for, and if the (diss)aggregated level of energy intensity allows for useful interpretations.

While results suggest that electricity customers are able to avoid transient inefficiency, we identify a larger share of underlying persistent inefficiency.
The following Section 2 provides a general overview of decoupling. Section 3 continues overviewing the methodology, defining the different models used in the analysis and discussing their econometric specifications. Section 4 describes the data used. We present the estimation results in Section 5 and finalize the paper with a conclusion in Section 6.

2. Background

The term decoupling is used in different and potentially confusing ways in the literature. Lowry and Makos (2010) consider revenue decoupling as the general, overall aim to tackle the throughput incentive. They argue that 'decoupling true up plans', 'lost revenue adjustment mechanisms (LRAMs)' as well as 'straight fixed variable pricing (SFV)' are the main three decoupling approaches. On the other hand, the EPA and the National Association of Regulatory Utility Commissioners (NARUC) argue that decoupling mechanisms (referring only to decoupling true up plans[2]) are one way to address the throughput incentive while LRAMs and SFV are different approaches and therefore should not be understood as decoupling[3] (EPA, 2007; NARUC, 2007). LRAMs allow utilities to recover efficiency-related revenue (sales) reductions, and thus target the misdirected incentive of avoiding efficiency improvements. By spreading the fixed costs of a utility equally to all its customers via a high fixed charge combined with low consumption based costs (price/kWh), SFV tariffs try to nivelise the sales maximization incentive of utilities directly. (EPA, 2007)
Following the definition of the EPA, decoupling can be further specified into different mechanisms that try to unlink total sales and revenue recovery. While certain approaches aim to preserve the recovery of lost margins, accounting for the simultaneous reduction of costs when sales decline, others offer limited true ups. The two most prominent decoupling approaches are total-revenue-caps and revenue-per-customer-caps. In both cases an allowed revenue is defined for a baseline year. A utility's unit sales price will be adjusted depending on whether the actual revenue has exceeded this threshold or has fallen short in the following year. In regular time increments the threshold has to be revised. Assuming there is a

[2]The count of 13 states and the District of Columbia also refers only to decoupling true up plans. If all three mechanisms would be considered, 26 states and the District of Columbia would have implemented decoupling policies in the electricity sector.

[3]We will control for all three policy mechanisms. However, we follow the definition of the government agencies and consider solely 'true up plans' as decoupling. For detailed descriptions and examples see NARUC (2007), EPA (2007) or Eto et al. (1997).

revenue-per-customer-cap, the actual revenue of a utility is compared periodically with its allowed revenue (while accounting for changes in the number of customers) and adjusted in the following period via modifying the unit sales price. A general side effect is that decoupling consolidates the revenue of companies confronted with a volatile energy demand (EPA, 2007).

It is important to state that decoupling was not initially developed to create energy efficiency, but to tackle the 'throughput incentive' (NARUC, 2007). Despite general concerns about implementation there is also a wide-ranging discussion regarding the actual effects of decoupling. The Electricity Consumers Resource Council (2007) and Brennan (2010) suggest that decoupling passes risks on to the consumers. When utilities' revenues are independent from sales, the utilities might lose incentives to ensure the security of supply (Brennan, 2010). Furthermore, secure revenues for utilities lead to less volatile costs for end consumers, reducing their incentives to invest in energy efficiency (Electricity Consumers Resource Council, 2007). Additionally, if some customers implement efficiency measures they can increase the electricity bills of those who do not invest in efficiency. This can put extensive pressure on low-income households, who cannot afford efficiency investments (NARUC, 2007). Brennan (2010) presents a thorough theoretical approach, modeling - inter alia - the behavior of electric utilities, that implement the decoupling mechanism. His results indicate that decoupled utilities will support energy efficiency investments and will supply demand-reducing information, but only if electricity sales prices are below their marginal costs of production. He concludes his analysis on 'decoupling in electric utilities' with the suspicion that the real reason why decoupling is implemented is that politicians want to avoid telling customers that they are responsible for increasing energy usage and climate pollution. Rather the responsibility for decreasing energy usage is pushed to the supply side.

Existing empirical work regarding revenue decoupling is limited. Focusing on Demand Side Management (DSM), Datta (2015) addresses the influence of decoupling on electric utility expenditures on energy efficiency. Using fixed effects regression models and utility specific data for the period of 2007 to 2011, he finds that utilities that have been decoupled spend on average four times as much money on energy efficiency than non-decoupled utilities. Kahn-Lang (2016) uses a game-theoretical approach to analyze potential connections between decoupling, residential energy consumption and utilities expenses in DSM. Using data for 218 non-governmental electricity utilities from $2001 - 2010$, his results show that decoupling has an indirect influence on decreased energy consumption as it promotes DSM spending and DSM efficiency. Finally, Brucal and Tarui (2018) investigate the effect of revenue decoupling on electricity prices and welfare. They utilize a panel data set covering 12 years and containing information about roughly 200 investor-owned electric utilities. By combining a propensity score matching approach with a difference-in-difference estimation, they find that soon after utilities implement a decoupling policy they charge significantly higher electricity prices. In fact, their results suggest that decoupling mechanisms are misused. While a decrease in sales leads to a rise of electricity prices an increase in sales does not lead to the reverse effect in the same order of magnitude. In terms of welfare, their results support the argument that revenue decoupling transfers economic risks from the supply side to the consumers (Brucal and Tarui, 2018).

Yet, none of these studies focus on the initial target of revenue decoupling to tackle the throughput incentive. To the best of our knowledge no one so far has implemented a SFA approach using a demand frontier function, as suggested by Filippini and Hunt (2015*a*), to investigate the influence of decoupling on electricity demand and efficiency. We find a significant negative effect of decoupling on electricity demand and thus contribute to the on-going discussion about the overall repercussions of decoupling. Furthermore, we detect the presence of persistent as well as transient (in)efficiency among electricity customers. However, the persistent inefficiency is substantially larger than the transient.

3. Empirical approach

Several approaches exist to date to investigate energy efficiency. As discussed before, energy intensity seems to be an inadequate proxy for energy efficiency. Metcalf (2008) therefore aims to decompose energy intensity into actual efficiency improvements and economic activity related changes of the intensity score. Similarly, Jimenez and Mercado (2014) decompose energy intensity through the Fisher Ideal Index into energy efficiency (real energy intensity) and economic activities. They find developments in energy efficiency to be the major driver of energy intensity change. However, activity changes also substantially contribute to energy intensity alterations. In addition to decomposing methods, a large share of literature aims to estimate efficiency through frontier analysis approaches. Those approaches are either non-parametric or parametric. Both define efficiency as a measure of comparison of a unit to a "best-practice" benchmark, while accounting for specific individual characteristics (Boyd, 2008).

The estimation of an input or distance demand frontier function, as well as the estimation of a cost or production frontier function rests upon the concept of productive efficiency, as originally introduced by Farrell (1957). Filippini and Hunt (2015*a*) define three major components that ensure that the production of an energy service is efficient, which is a necessity for economic success. These are the minimization of inputs to produce a given output-level, the selection of the best (in financial terms) combination of input factors, and the utilization of the lowest cost technology.

Concepts to investigate efficiency by using non−parametric approaches and linear programming are generally referred to as Data Envelopment Analysis (DEA). A specification to investigate energy efficiency using DEA models was developed by Zhou and Ang (2008). Their approach is to measure economy-wide energy efficiency performance based on a joint production framework. Advantages of such models are that they can be used to analyze rather small datasets. Furthermore, non-parametric approaches don't impose a specific form of the production (or cost) function. However, as those approaches are unable to account for unobserved heterogeneity, the present study follows the argumentation of Filippini and Hunt (2015*a*) to use a parametric method, namely SFA.

Using SFA to estimate a Shepard energy distance function, Lin and Du (2013) use a meta-frontier concept to evaluate energy efficiency developments of 30 administrative regions in China. Buck and Young (2007) investigate the energy efficiency in the Canadian commercial building sector using the SFA approach as presented by Aigner et al. (1977). Specifically, they calculate the ratio of the actually used energy in a building to the hypothetical optimal level of energy usage. A similar approach was conducted by Boyd (2008), who interprets the stochastic frontier function as an input-distance function to estimate efficiencies at plant level.

While the proposed frontier function by Aigner et al. (1977) is based on the neoclassical theory of production, Filippini and Hunt (2011) utilize their approach to estimate an aggregated energy demand function, assuming an underlying 'production process'. By analyzing the energy efficiency development of 29 OECD countries via the joint use of energy demand modeling and a stochastic frontier approach they introduce a new tool to estimate energy demand and efficiency. Furthermore, their results support the proposition that energy intensity is not a suitable proxy for energy efficiency.

Following this concept, Filippini and Hunt (2012) estimate a Pooled Model, a Random Effects Model (REM) and a Mundlak corrected version of a REM (MREM) to investigate US residential energy demand and efficiency in 48 U.S. states. Filippini et al. (2014) pick up the SFA approach developed by Filippini and Hunt (2011, 2012) to estimate a residential frontier energy demand function, in order to evaluate the effects of energy-efficiency policy

measures within the EU.
However, none of the listed studies differentiates between persistent and transient efficiency. There are in fact only few studies trying to separate both levels of efficiency. Tsionas and Kumbhakar (2012) present a SFA model that allows this differentiation. They develop a generalized true random effects model, which differentiates time-invariant effects into a persistent (time-invariant) inefficiency effect and a random firm effect (to capture heterogeneity). Additionally, they introduce a time-variant inefficiency component, leading to the construction of a 4-component error-term. This model thus differentiates between persistent and transient efficiency and captures random firm effects and general noise. Additional studies that are based on this theoretical approach are, amongst others, Kumbhakar et al. (2012) and Colombi et al. (2014).
A different, and for our study, more applicable approach was developed by Filippini and Hunt (2015*b*), which is the starting point for the analysis presented here. While discussing three types of parametric approaches, namely the estimation of an input requirement function (as proposed by Kumbhakar and Hjalmarsson (1995)), a Shepard energy input distance function (e.g. conducted by Lin and Du (2013)) and an input demand frontier function (as developed by Filippini and Hunt (2011)), they suggest the use of the latter. Despite having some limitations the key advantage of this function is that it takes allocative as well as technical efficiency into account (Filippini and Hunt, 2015*a*)[4].
Furthermore, Filippini and Hunt (2015*a*) suggest estimating two models separately: A Mundlak adjusted random effects (MREM) model to estimate the persistent part of the level of energy inefficiency, and a true random effects model (TREM) to estimate the transient part. One of the classic SFA random effects model (REM), which was developed by Pitt and Lee (1981) (in the following PL-model), is based on the assumption that the inefficiency term u_i is constant over time. Therefore it only captures the persistent level of energy efficiency.
Persistent levels of energy efficiency are entirely caught by a utility-specific constant term when a TREM is used. Thus, the estimation of a TREM creates efficiency values that are transient, as they vary over time. [5] One key-feature of the TREM, which was developed by Greene (2005*a*,*b*) is the ability to differentiate between transient efficiency and unobserved heterogeneity. However, Greene (2005*a*) also argues that the PL-model is unsuitable to

[4]For a detailed discussion regarding the different model specifications and empirical examples see Filippini and Hunt (2015*b*).

[5]Nevertheless, the exclusion of persistent energy efficiency must be taken into account when interpreting the estimated efficiencies.

identify persistent inefficiency. Especially, as the time invariant component will be captured not only in the inefficiency term but also in the constant α, which are "indecomposable". Investigating the Swiss railway sector, Filippini and Greene (2015) take up this argument to explain the weak correlation between the estimated efficiency scores based on a PL-model and the persistent efficiency values estimated using the Generalized True Random Effects Model (GTREM), which they present in this paper. They extend the TREM, which was developed by Greene (2005*a*,*b*), by a second disturbance parameter. Therefore, the GTREM contains one time invariant and one time variant efficiency component. The reduced complexity of the maximization of the underlying log likelyhood function is an essential advantage compared to the earlier mentioned model of Tsionas and Kumbhakar (2012). This is achieved by using a simulation approach and by utilizing the work of Butler and Moffitt (1982)[6]. Empirical applications of this model are done by Filippini et al. (2017), who estimate a translog cost function using data on hydro-powered electricity production in Switzerland, and Blasch et al. (2017), who utilize unique survey data to estimate transient and persistent efficiency of Swiss residential households.

We utilize the GTREM as third model, to correctly identify the persistent and transient efficiency. In line with Filippini and Greene (2015) we additionally estimate a MREM (PL-model) and the TREM developed by Greene (2005*a*,*b*), as originally suggested by Filippini and Hunt (2015*a*). However, we include Mundlak corrections in all estimations. Table 1 presents the characteristics of the three models.

Table 1: Model characteristics

Model I - REM	Model II - TREM	Model III - GTREM
$lnE_{it} = \alpha + \beta' x_{it} + v_{it} + u_i,$	$lnE_{it} = \alpha_i + \beta' x_{it} + v_{it} + u_{it},$	$lnE_{it} = \alpha_i + \beta' x_{it} + v_{it} + u_{it},$
$v_{it} \sim N[0, \sigma_v^2],$	$\alpha_i = \alpha + w_i, w_i \sim N[0, \sigma_w^2],$	$\alpha_i = \alpha + (w_i - h_i)$
$u_i = \lvert U_i \rvert, U_i \sim N[0, \sigma_u^2],$	$v_{it} \sim N[0, \sigma_v^2],$	$w_i \sim N[0, \sigma_w^2],$
$\epsilon_{it} = v_{it} + u_i$	$u_{it} = \lvert U_{it} \rvert, U_{it} \sim N[0, \sigma_u^2],$	$h_i = \lvert H_i \rvert, H_i \sim N[0, \sigma_h^2],$
	$\epsilon_{it} = v_{it} + u_{it}$	$v_{it} \sim N[0, \sigma_v^2],$
		$u_{it} = \lvert U_{it} \rvert, U_{it} \sim N[0, \sigma_u^2],$
		$\epsilon_{it} = v_{it} + u_{it}$

Within the framework of the REM v_{it} captures random noise and u_i is non-positive ($u \leq 0$) and represents time-invariant (persistent) inefficiency. Regarding the TREM, the constant term additionally contains w_i, which is a random component within the random-effects framework that is independent and identically distributed (iid). The error term

[6]For a detailed description of the model derivation and an empirical example see Filippini and Greene (2015).

contains two components: v_{it} to capture noise and the inefficiency term u_{it}. Based on the assumption that u_{it} is non-normal distributed (namely: exponential, half-normal or truncated-normal) and fulfills the iid-conditions, the inefficiency must vary over time (Filippini and Greene, 2015).

By displaying the GTREM it becomes clear that the key difference to the TREM lies within the integration of the second disturbance parameter h_i, which captures time-invariant inefficiency. The random components w_i,v_{it}, and u_{it} are defined as in the TREM and the additional h_i is half-normal distributed. In terms of estimation complexity, it is essential to know that these four components add up to only two disturbance parameters and not four.

On the basis of the assumed underlying aggregated energy demand function (following Filippini and Hunt (2011, 2015*b*,*a*), we estimate a stochastic frontier function that is constructed as follows:

$$\begin{aligned} lnE_{it} = \alpha &+ \beta_p lnP_{it} + \beta_{gdp} lnGDP_{it} + \beta_{cus} lnCUS_{it} + \beta_{hdd} lnHDD_{it} + \beta_{cdd} lnCDD_{it} \\ &+ \beta_{shi} SHI_{it} + \beta_{shs} SHS_{it} + \beta_{shp} SHP_{it} + \beta_{true} TRUE_{it} + \beta_{lram} LRAM_{it} \\ &+ \beta_{sfv} SFV_{it} + \beta_{eers} EERS_{it} + \beta_t T + \epsilon \end{aligned} \tag{1}$$

where lnE_{it} represents the logarithmised energy consumption in the sphere of utility 'i' in period 't' (sold electricity), lnP_{it} the logarithm of the average state-level electricity price[7], $lnGDP_{it}$ the logarithm of GDP per capita, and $lnCUS_{it}$ the logarithm of the total number of customers. $lnHDD_{it}$ and $lnCDD_{it}$ are the logarithmised values of the heating degree days and cooling degree days respectively. SHI_{it}, SHS_{it} and SHP_{it} are the industrial electricity consumption, the commercial and the residential shares respectively. $TRUE_{it}$, $LRAM_{it}$, SFV_{it} and $EERS_{it}$ are qualitative variables that indicate whether a utility 'i' has implemented a true up plan, LRAM, SFV or EERS respectively, in period 't'. Finally T represents year fixed effects to account for technology changes over time and ϵ represents the error term, whose specific design depends on the estimated model.

With respect to the REM and the TREM, to measure the efficiency of the customers of an electricity utility, the energy efficiency score EE_{it} is defined as follows:

$$EE_{it} = \frac{E^F_{it}}{E_{it}} = exp(-\hat{u}_{it}) \tag{2}$$

[7] We are aware of the potential endogeneity that comes along with the usage of a price variable. However, we assume non-monopolistic markets, in which a utility can not freely set its selling price and thus changes of the sold electricity of a single utility should not be able to influence the average state-level electricity price.

where E_{it}^F represents the frontier estimated in (1), which thus represents the (efficient) minimum demand of the customers of utility 'i' in period 't', and E_{it} represents the actual energy demand of the 'i's utilities customers in period 't'. Therefore, each utility that lies on the frontier gets an EE-score (EE_{it}) equal to one, which implies that they are 100% efficient. EE-scores of utilities apart from the frontier are between zero and one.

4. Data description

The basis of our analysis is an unbalanced data set containing information for all U.S. states[8] and the District of Columbia from 2001 until 2015. The dataset covers 209 electric utilities with an average 13.1 years within the considered period (N = 2739). Table 2 presents descriptive statistics of the data.

Table 2: Descriptive Statistics

	Unit	Mean	Std. Dev.	Minimum	Maximum
Total Sales	GWh	10,634.23	14,881.25	9.53	87,160.37
State-level Electricity Price	US ct/kWh	9.01	3.38	4.24	34.04
Real GDP per Person (2009)	US $	46,632.74	11,575.73	28,856.00	170,687.00
Share of Industrial Customers	%	26.01	17.60	0.00	75.68
Share of Residential Customers	%	39.76	13.94	10.70	81.59
Share of Commercial Customers	%	33.82	12.08	6.94	85.14
Total Number of Customers	-	443,876.80	700,692.10	1001	5,268,369
Heating Degree Days	-	5,154.82	2,167.50	0	11,702
Cooling Degree Days	-	1,191.07	909.16	2	4,904
EERS	-	0.32	0.47	0	1
True Up Plan	-	0.06	0.24	0	1
LRAM	-	0.08	0.28	0	1
SFV	-	0.01	0.11	0	1

Note: HDD (CDD) are measured as the absolute year-sum of the daily differences between the average day-temperature and the set base temperature of 65 F° if the difference is negative (positive).

The US Energy Information Administration (EIA) grants access to utility-specific data regarding total sales, consumption shares (SHI, SHP and SHS) as well as the total number of customers. They provide detailed data files based on the Form EIA-861, which comprehensively presents information raised in the 'Annual Electric Power Industry Report' online on their web-page. The EIA also provides data on the average state-level electricity price

[8]The state of Nebraska is later excluded from the sample, as no electric utility is investor-owned in this state.

within the data files based on EIA-861. We correct those average state-level prices for inflation to the base year 2009, by using the price index for private consumption expenditures, which is provided by the Bureau of Economic Analysis (BEA) (U.S. Department of Commerce). Furthermore, the BEA provides state-level real GDP per capita values, chained to the same base year of 2009. Information on HDD and CDD is jointly provided online by the U.S. National Weather Service, which is part of the National Oceanic and Atmospheric Administration (U.S. Department of Commerce). Information on EERS is provided by the American Council for Energy-Efficient Economy (ACEEE, 2017). The EEI (2015) provides year-specific information on whether a utility implemented a decoupling true up plan or a LRAM. They also present information, although not always year-specific, on implemented SFV mechanisms. Figure 1 illustrates the deployment development of the mechanisms aiming to tackle the 'throughput incentive' across the observed period.

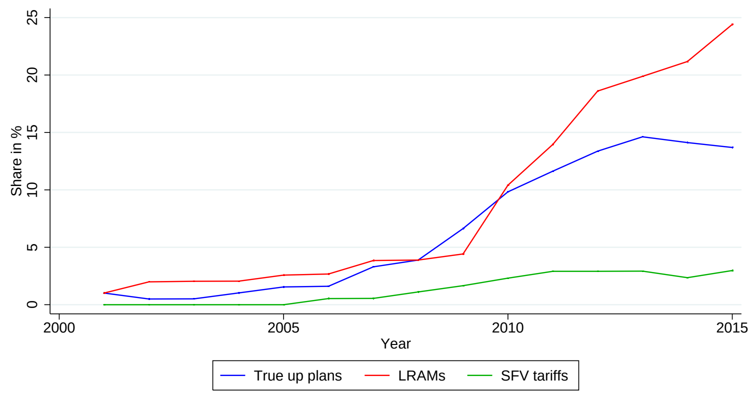

Figure 1: Development of policy mechanisms

In the year 2001 very few electric utilities had any of the three policies implemented. Overall, true up plans and LRAMs experience a rise in popularity with strongly growing deployment from 2008 on. In 2015 almost 15% of electric utilities implemented a true up plan and around 25% a LRAM. The share of utilities who implemented a SFV-tariff remains below 5% throughout our sample.

Our dataset is restricted in the following ways. For utilities that start or end their operation within our investigated period, we delete their last (first) year, if an outstanding difference in electricity sells to the previous (following) year indicates that the utility did not end (start) its operations at the end (beginning) of the considered year. Additionally, we drop observations with less than 1,000 customers, with 0 kWh sold, and with an average of more than 100,000 kWh/customer, as those utilities are not comparable to the rest of our sample[9]. For a few utilities we have ambiguous information regarding the implementation of a SFV tariff, thus we exclude them too. Finally, we eliminate four utilities, who have at least one gap, e.g. a missing year, within their operation period.

5. Estimation results

The estimation results are presented in Table 3. The computation was done using the software LIMDEP/NLOGIT, where commands for all our models are available. All three models show quantitatively and qualitatively very similar coefficients. Therefore, we focus our interpretation on the MGTREM.

Table 3: Estimation results

	Model I (MREM)		Model II (MTREM)		Model III (MGTREM)	
Variable	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error
Constant	1.812	(3.657)	0.849***	(0.116)	0.450***	(0.111)
St.-lvl elec. pr. (log)	−0.100***	(0.007)	−0.103***	(0.006)	−0.099***	(0.006)
GDP per capita (log)	0.242***	(0.009)	0.236***	(0.006)	0.245***	(0.008)
No. of customers (log)	0.962***	(0.004)	0.975***	(0.003)	0.977***	(0.003)
HDD (log)	0.129***	(0.015)	0.129***	(0.012)	0.130***	(0.012)
CDD (log)	0.040***	(0.005)	0.043***	(0.004)	0.046***	(0.003)
SHI	0.286***	(0.056)	0.260***	(0.050)	0.258***	(0.050)
SHS	0.000	(0.055)	−0.030	(0.050)	−0.033	(0.049)
SHP	−1.837***	(0.056)	−1.841***	(0.051)	−1.850***	(0.051)
True Up Plan	−0.037***	(0.003)	−0.038***	(0.003)	−0.038***	(0.003)
LRAM	0.002	(0.003)	0.003	(0.002)	0.003	(0.002)
SFV	0.043***	(0.005)	0.044***	(0.005)	0.048***	(0.005)
EERS	−0.001	(0.002)	−0.001	(0.002)	−0.001	(0.002)
Year fixed effects	yes		yes		yes	
Observations (N)	2739		2739		2739	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Both, MREM and MGTREM, are based on 1000 Halton draws.
Note: The Mundlak corrections are excluded from the table to safe space.

[9] The estimated coefficients are qualitatively and quantitatively robust to variations of those thresholds throughout all models. However, the minimum values of the estimated efficiencies are implausibly small, if we include utilities with less customers or with a higher sells/customer-ratio.

The estimated mean parameter for the decoupling true up plan is highly statistically significant and negative. The parameter indicates that revenue decoupling true up plans are associated with a 3.752%[10] reduction of the electricity consumption. The findings of Brucal and Tarui (2018) that decoupled utilities charge higher electricity prices than non-decoupled ones, are supported by these findings. If we acknowledge the mechanism of decoupling, which only allows utilities to increase the electricity prices if sales are less than expected, the significant associated reduction of electricity sales should lead to an increase in kWh prices.
Furthermore, we can not identify any correlation between LRAMs and energy consumption, and even a highly statistically significant positive correlation between SFV tariffs and energy consumption. Given the design of SFV tariffs (high fixed costs and almost no variable costs), this finding is intuitive and suggests that those tariffs fail to tackle the throughput incentive by ignoring corresponding incentives for the demand-side.
It seems straight forward that the average state-level electricity price has a significant negative relation with the electricity consumption. Also the results regarding the GDP and the total number of customers are as expected. An increase of customers by 1% is associated with an estimated increase of electricity consumption by also around 1% (0.962% - MREM, 0.975% - MTREM, and 0.977% - MGTREM). Similarly, GDP per capita has a highly statistically significant positive relation with the electricity consumption. The positive and highly statistically significant correlations between HDDs/CDDs and electricity consumption are also intuitive. The increased use of air conditioning or heating should increase the total electricity consumption. We further identify that the share of industrial consumption (SHI) is statistically significant and positively associated with electricity consumption, while the residential share has a statistically significant negative relation. The estimated parameter regarding the EERS is neither statistically significantly different from zero nor economically relevant. By controlling for year fixed effects we account for technological changes over time. Acknowledging the findings of Datta (2015) and Kahn-Lang (2016), we control for DSM-spending as a robustness check. However, the relevant data availability is limited, which not only restricts our sample to the period of 2001 - 2012, but also excludes some utilities. Throughout all models, we can not find a significant correlation between the DSM-spending of a utility and its electricity sales. Apart from the commercial sector (SHS), the estimated parameters remain qualitatively the same. Quantitative changes are most-likely caused by the sample restrictions.

[10] As discussed by Giles (1982) to be the preferred method, we follow the approach of Kennedy (1981) and estimate the effect size of the dummy variable using $e^{\beta_1 - 1/2V(\beta_1)} - 1$ ($\hat{\beta}_{true} = -0.04278$; $\hat{V}(\beta_{true}) = 0.00248$).

Furthermore, by using Equation 2 and utilizing the approach presented in Filippini and Greene (2015) we obtain efficiency estimates shown in Table 4. While the transient efficiency scores based on the MTREM an MGTREM are highly correlated (0.864) the estimated persistent efficiency values based on the MREM and the MGTREM are only moderately correlated (0.127). The rather small correlation value indicates biased estimates with respect to the MREM. This is in line with the argumentation of Greene (2005*a*) that long-run efficiency estimates based on a PL-model are potentially biased. This biased can be caused by any underlying unobserved heterogeneity, which the REM identifies as inefficiency (Blasch et al., 2017). This could also explain why the efficiency scores of the REM are substantially smaller than those of the GTREM. For this reason, we do not further interpret the long-run efficiency based on the MREM.

Table 4: Summary of efficiency estimates

	Mean	Std. Dev.	Minimum	Maximum
Long-run EE (MREM)	0.6191	0.1562	0.2890	0.9931
Short-run EE (MTREM)	0.9598	0.0232	0.7445	0.9954
Long-run EE (MGTREM)	0.8314	0.0010	0.8279	0.8373
Short-run EE (MGTREM)	0.9560	0.0083	0.7924	0.9932

In line with the results of Filippini and Hunt (2015*a*), who estimated energy efficiency in the U.S. on a state-level basis, we find a larger share of persistent than transient energy inefficiency. We identify a low level of variation in the transient efficiency around the average of 96%. Furthermore, the estimated persistent efficiency indicates a sector-wide inefficiency of almost 17%. Aggregated, the GTREM predicts an average electricity inefficiency of around 20% between 2001 and 2015.

6. Conclusions

Different methods exist to estimate energy efficiency. Common are decomposition techniques and the estimation of efficiency via parametric or non-parametric frontier approaches. However, most previous studies do not differentiate between persistent (time-constant) and transient (time-varying) efficiency. We pursue the approach proposed by Filippini and Hunt (2015*a*,*b*) to estimated an energy demand function using SFA. In addition to their suggested models, we utilize the GTREM developed by Filippini and Greene (2015), which allows the differentiation between persistent and transient efficiency. By taking up the discussion about biased efficiency estimates based on a PL-model (Greene (2005*a*); Filippini and Greene (2015)), we integrate our empirical analysis into this strand of literature.

The estimation results allow the conclusion that persistent inefficiency is substantially larger in the electric utility sector in the U.S. than transient inefficiency. Customers of investor-owned electric utilities seem able to avoid transient inefficiency quite well. However, results indicate an underlying persistent inefficiency across the electricity sector, which should be a concern for electricity regulators.

Investigating the impact of implemented decoupling mechanisms on investor-owned electricity utilities' customers, we find a significant negative relationship between revenue decoupling true up plans and the electricity consumption of the effected customers. Furthermore, we can not find any statistically significant correlation with respect to LRAMs and even a positive one regarding SFV tariffs. Based on our results, we can conclude that solely decoupling true up plans are actually associated with a reduction of the sold electricity of electric utilities, indicating their ability to tackle the throughput incentive.

References

ACEEE (2017), 'State energy efficiency resource standards (EERS) January 2017'. American Council for an Energy-Efficient Economy.

Aigner, D., Lovell, C. and Schmidt, P. (1977), 'Formulation and estimation of stochastic frontier production function models', *Journal of Econometrics* **6**(1), 21–37.

Ang, B., Mu, A. and Zhou, P. (2010), 'Accounting frameworks for tracking energy efficiency trends', *Energy Economics* **32**(5), 1209–1219.

Ang, J. B. (2007), 'CO2 emissions, energy consumption, and output in france', *Energy Policy* **35**(10), 4772–4778.

Apergis, N. and Payne, J. E. (2010), 'The emissions, energy consumption, and growth nexus: Evidence from the commonwealth of independent states', *Energy Policy* **38**(1), 650–655.

Blasch, J., Boogen, N., Filippini, M. and Kumar, N. (2017), 'Explaining electricity demand and the role of energy and investment literacy on end-use efficiency of swiss households', *Energy Economics* **68**, 89–102.

Boyd, G. A. (2008), 'Estimating plant level energy efficiency with a stochastic frontier', *The Energy Journal* **29**(2).

Boyd, G. A. and Roop, J. M. (2004), 'A note on the fisher ideal index decomposition for structural change in energy intensity', *The Energy Journal* **25**(1).

Brennan, T. J. (2010), 'Decoupling in electric utilities', *Journal of Regulatory Economics* **38**(1), 49–69.

Brucal, A. and Tarui, N. (2018), 'Revenue decoupling for electric utilities: Impacts on prices and welfare', *Working Paper* .

Buck, J. and Young, D. (2007), 'The potential for energy efficiency gains in the canadian commercial building sector: A stochastic frontier study', *Energy* **32**(9), 1769–1780.

Butler, J. S. and Moffitt, R. (1982), 'A computationally efficient quadrature procedure for the one-factor multinomial probit model', *Econometrica* **50**(3), 761.

Colombi, R., Kumbhakar, S. C., Martini, G. and Vittadini, G. (2014), 'Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency', *Journal of Productivity Analysis* **42**(2), 123–136.

Datta, S. (2015), 'Decoupling and demand side management'. Presented at the European Association of Environmental and Resource Economists 21st Annual Conference.

EEI (2015), *Alternative Regulation for Emerging Utility Challenges: 2015 Update*, Edison Electric Institute, prepared by: Pacific Economics Group Research LLC.

Electricity Consumers Resource Council (2007), 'Isos, rtos ineffective, elcon testifies at ferc conference'.

EPA (2007), 'Aligning utility incentives with investment in energy efficiency'. U.S. Environmental Protection Agency.

EPA (2008), 'National action plan for energy efficiency vision for 2025: A framework for change - a resource of the national action plan for energy efficiency'. U.S. Environmental Protection Agency.

EPA (2017), 'Inventory of u.s. greenhouse gas emissions and sinks 1990 - 2015'. U.S. Environmental Protection Agency.

Eto, J., Stoft, S. and Belden, T. (1997), 'The theory and practice of decoupling utility revenues from sales', *Utilities Policy* **6**(1), 43–55.

Farrell, M. J. (1957), 'The measurement of productive efficiency', *Journal of the Royal Statistical Society. Series A (General)* **120**(3), 253.

Filippini, M., Geissmann, T. and Greene, W. H. (2017), 'Persistent and transient cost efficiency—an application to the swiss hydropower sector', *Journal of Productivity Analysis* **49**(1), 65–77.

Filippini, M. and Greene, W. (2015), 'Persistent and transient productive inefficiency: a maximum simulated likelihood approach', *Journal of Productivity Analysis* **45**(2), 187–196.

Filippini, M. and Hunt, L. C. (2011), 'Energy demand and energy efficiency in the OECD countries: A stochastic demand frontier approach', *The Energy Journal* **32**(2).

Filippini, M. and Hunt, L. C. (2012), 'US residential energy demand and energy efficiency: A stochastic demand frontier approach', *Energy Economics* **34**(5), 1484–1491.

Filippini, M. and Hunt, L. C. (2015*a*), ' Measuring persistent and transient energy efficiency in the US', *Energy Efficiency* **9**(3), 663–675.

Filippini, M. and Hunt, L. C. (2015*b*), ' Measurement of energy efficiency based on economic foundations', *Energy Economics* **52**, S5–S16.

Filippini, M., Hunt, L. C. and Zorić, J. (2014), 'Impact of energy policy instruments on the estimated level of underlying energy efficiency in the EU residential sector', *Energy Policy* **69**, 73–81.

Giles, D. E. (1982), 'The interpretation of dummy variables in semilogarithmic equations', *Economics Letters* **10**(1-2), 77–79.

Greene, W. (2005*a*), 'Fixed and random effects in stochastic frontier models', *Journal of Productivity Analysis* **23**(1), 7–32.

Greene, W. (2005*b*), 'Reconsidering heterogeneity in panel data estimators of the stochastic frontier model', *Journal of Econometrics* **126**(2), 269–303.

IEA (2009), 'Progress with implementing energy efficiency policies in the G8'. International Energy Agency.

Jimenez, R. and Mercado, J. (2014), 'Energy intensity: A decomposition and counterfactual exercise for latin american countries', *Energy Economics* **42**, 161–171.

Kahn-Lang, J. (2016), 'The effects of electric utility decoupling on energy efficiency', *The Energy Journal* **37**(4).

Kennedy, P. (1981), 'Estimation with correctly interpreted dummy variables in semilogarithmic equations [the interpretation of dummy variables in semilogarithmic equations]', *American Economic Review* **71**(4). **URL:** *https://EconPapers.repec.org/RePEc:aea:aecrev:v:71:y:1981:i:4:p:801*

Kumbhakar, S. C. and Hjalmarsson, L. (1995), 'Labour-use efficiency in swedish social insurance offices', *Journal of Applied Econometrics* **10**(1), 33–47.

Kumbhakar, S. C., Lien, G. and Hardaker, J. B. (2012), 'Technical efficiency in competing panel data models: a study of norwegian grain farming', *Journal of Productivity Analysis* **41**(2), 321–337.

Lin, B. and Du, K. (2013), 'Technology gap and china's regional energy efficiency: A parametric metafrontier approach', *Energy Economics* **40**, 529–536.

Lowry, M. N. and Makos, M. (2010), 'Review of distriubution revenue decoupling mechanisms'. Pacific economics group research LLC.

Metcalf, G. E. (2008), 'An empirical analysis of energy intensity and its determinants at the state level', *The Energy Journal* **29**(3).

NARUC (2007), *Decoupling For Electric & Gas Utilities: Frequently Asked Questions (FAQ)*, The National Association of Regulatory Utility Commissioners (NARUC).

OECD (2012), *OECD Environmental Outlook to 2050*, OECD Publishing.

Pitt, M. M. and Lee, L.-F. (1981), 'The measurement and sources of technical inefficiency in the indonesian weaving industry', *Journal of Development Economics* **9**(1), 43–64.

Tsionas, E. G. and Kumbhakar, S. C. (2012), 'FIRM HETEROGENEITY, PERSISTENT AND TRANSIENT TECHNICAL INEFFICIENCY: A GENERALIZED TRUE RANDOM-EFFECTS model', *Journal of Applied Econometrics* **29**(1), 110–132.

Victor, D. G., Akimoto, K., Kaya, Y., Yamaguchi, M., Cullenward, D. and Hepburn, C. (2017), 'Prove paris was more than paper promises', *Nature* **548**(7665), 25–27.

Voigt, S., Cian, E. D., Schymura, M. and Verdolini, E. (2014), 'Energy intensity developments in 40 major economies: Structural change or technology improvement?', *Energy Economics* **41**, 47–62.

Zhou, P. and Ang, B. (2008), 'Linear programming models for measuring economy-wide energy efficiency performance', *Energy Policy* **36**(8), 2911–2916.