11-2017

# Econometric Estimation of Groundwater Depth Change for the High Plains Aquifer

Jonathan R. Sims
*University of Nebraska-Lincoln*, jonsims06@gmail.com

# ECONOMETRIC ESTIMATION OF GROUNDWATER DEPTH CHANGE FOR THE HIGH PLAINS AQUIFER

by

Jonathan R. Sims

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Agricultural Economics

Under the Supervision of Professor Lilyan E. Fulginiti

Lincoln, Nebraska

November 2017

ECONOMETRIC ESTIMATION OF GROUNDWATER DEPTH CHANGE FOR THE HIGH PLAINS AQUIFER

Jonathan R. Sims, M.S.

University of Nebraska, 2017

Advisor: Lilyan E. Fulginiti

This article presents a new method for estimating changes in depth to groundwater at a yearly, county level and incorporates these estimates as the dependent variable of econometric models for the High Plains aquifer. The High Plains (Ogallala) aquifer underlies eight states in the central United States and is the primary source of irrigation water for this large food producing region. The stock of groundwater is a finite, non-renewable resource with minimal recharge in most areas. Many fields of study, including hydrology and agricultural economics, are interested in depth to groundwater changes because they serve as a proxy for estimating groundwater stock changes. Economic data exist at the yearly, county level, but there are currently no yearly estimates for depth to groundwater changes making it difficult to reliably utilize economic optimization and production models that depend on groundwater data. Including the new estimates generated in this study as the dependent variable with climate, recharge, and irrigation as independent variables in panel econometric models (Pooled OLS, Random Effects, and Fixed Effects) with counties as the individuals produced statistically significant results. Further, models were found which consistently performed best when comparing coefficients and predicted values with outside estimates from hydrology studies.

# ACKNOWLEDGEMENTS

I would like to thank my parents, Bob and Jean. They walked with me through the most trying times, encouraged me to find my own way forward, and believed in me when I did not.

I am grateful and fortunate to have been advised by professors Lilyan Fulginiti, Richard Perrin, and Karina Schoengold at UNL. They not only motivated me to pursue this project from the start, but also made it possible to finish through continual patience, support, and belief. My gratitude for them is as much personal as it is professional.

I could not have completed this thesis work without the significant financial support of the UNL Agricultural Economics Department. Its financial investments in students is apparent from the strength of graduates and quality of research being produced.

Thank you to professor Allen Dalton at Boise State University for drawing me to the field of economics. I try to take his enthusiasm and clarity of mind with me into every intellectual pursuit.

Thank you to Max, Carl, and Justin for their friendship and direction. They were instrumental in helping me cross the finish line with some sanity left to spare.

**Table of Contents**

1. Background……………………………………………………………………..……1

1.1 Introduction……………………………………………………..………..1

1.2 Motivation…………………………………………………………….…….2

2. Estimating Changes in Depth to Groundwater………………….……………………..3

2.1 Motivation……………………………………………………..…………3

2.2 A New Estimation Method……………………..……………………………4

2.3 Comparison with Existing Estimations……………..…….……………………6

3. Econometric Estimation……………………………….………..…………………8

3.1 Variables and Model Review…………………..……………………………8

3.2 Model Structure and Results…………………………………………………13

3.3 Inferences and Conclusions……………………….………………………15

References……………………………………………………………..………………20

Appendix……………………………………………………………………………22

## 1. Background

### *1.1 Introduction*

Groundwater-fed irrigation has been continually increasing over the past few decades, notably in India, China, and the United States (Scanlon et al., 2010). In the United States, groundwater withdrawals for irrigation account for around 30% of total U.S. groundwater use (Qi et al., 2002; Scanlon et al., 2010). The High Plains Aquifer (HPA) underlies about 174,000 square miles and eight mid-western States (WY, SD, NE, CO, KS, OK, TX, and NM). This is one of the most important food producing regions in the world (Steward et al., 2013).

As of the early-1980s—the beginning of this study's data set—about 170,000 wells were pumping water to irrigate about 13 million acres in the HPA (Gutentag et al., 1984). The 13 million irrigated acres provide high yields of corn, wheat, grain sorghum, soybeans, and cotton (Kromm and White, 1986). While so much of the economy and society depend upon this crucial resource, it receives little to no recharge relative to withdraws making it a finite resource for almost[1] every county which draws upon it. In parts of the [central and southern] High Plains, annual pumpage of 2 to 100 times greater than annual recharge has caused large water-level declines (Gutentag et al., 1984). Ground-water withdrawals from the HPA for irrigation increased from 4 to 19 million acre-feet from 1949 to 1974. Groundwater withdrawals for irrigation in 1980, 1985, 1990, and 1995 were 4 to 18 percent less than withdrawals for irrigation in 1974 (Heimes and Luckey, 1982). However, groundwater withdrawals from the aquifer for irrigation

1. some of the counties in western Nebraska—Especially the Sandhills in the northwest part of the state—receive large amounts of recharge and have two to three times the average saturated thickness of the HPA

were 21 million acre-feet in 2000 (Maupin and Barber, 2005), the largest for any 5-year period since the previous high-mark of 1974.

*1.2 Motivation*

Significant economic, environmental, and demographic changes could follow a reduction in irrigation as the groundwater continues to be depleted (Kromm and White, 1986). Further, although higher energy prices and costs of pumping may cause some short term effects on crop switching and lower returns, they are projected not to generate long term effects on abandoning irrigated production while water remains available (High Plains Study Council, 1982). Thus, it is important to have some estimates and knowledge about where and how quickly irrigation is depleting the stock of water in the HPA, as well as the interaction of climate on that relationship.

This study is entirely motivated by the problem of freshwater scarcity in the High Plains Region and its large-scale impact on the region's economies. The study's focused aim, then, is to contribute new estimation and econometric analysis methods for groundwater level changes at the yearly and county levels to be used in economic production and optimization models. Most economic data on inputs, outputs, and prices are at the county and yearly level. However, the majority of the hydrology literature estimates either large time scale changes (i.e. from pre-develpment to the 1990s or 2000s, or some five-year increments since 1990) (McGuire 2012a) or with high time resolution (continuous) measurements at monitoring wells. Multi-year and decade-long estimates of groundwater level changes do not provide enough information for economic production and optimization models, while high resolution monitoring wells are not numerous enough to give a reliable picture of county-level changes needed for the economic models this study hopes to be the foundation of.

To look at groundwater changes in a way more conducive for economic models, yearly, county-level changes are needed for as many of the counties over the aquifer as possible and an econometric model which can estimate or forecast groundwater-level changes using existing climate and irrigation data available. The first section of the paper describes a new method for estimating depth to groundwater changes at the county and yearly levels. The second section presents the econometric models, variables used, and interpretations of the model results where the groundwater-level change estimates from section one are implemented as the dependent variable in each model.

## 2. Estimating Changes in Depth to Groundwater

### *2.1 Motivation*

The reason for choosing changes in the depth to groundwater as the dependent variable is a matter of convenience since we would actually like to know the changes in water in storage under a county, that is, purely the change in the volume of water. However, estimating the change of water in storage requires additional estimates of the specific yield under a county and estimates of saturated thickness changes which require additional estimates of depth to the bottom of the aquifer. Changes in depth to water serve as a good proxy to water in storage since essentially all of the bottom of the aquifer is impermeable bedrock (Scanlon et al., 2010) and there is a vast collection of over 1 million depth to water measurements collected by the USGS over the time period of this study (1980-2010) and far fewer for depth to bedrock.

*2.2 A New Estimation Method*

Depth to groundwater measurements for individual wells from 1980-2010 were downloaded from the NWIS database. These contain all USGS well measurements in the NWIS database. Each data point included: site (well) ID number, depth to groundwater, factor—if any—affecting measurement accuracy, and measurement date. Depth to water measurements in June through September were removed to reduce unwanted variability from the effect of intra-seasonal depth changes due to cones of depression in the water table forming around wells from pumping before, during, and after the growing season. Figure 1 shows an example of the effect of intra-seasonal pumping on water table elevation across three years for four different wells. These were measured using continuous monitoring wells from the NWIS database[2] for visual and presentation purposes and were not used in the dataset.

Multiple steps were taken to remove erroneous depth to water measurements in the dataset prior to estimating county-level depth to groundwater changes. Depth to water measurements in the data which exceeded 2000 feet were removed as there were multiple erroneous measurements of up to 1 million feet. The dataset included factors which might affect measurement accuracy—including nearby pumping and recent pumping at the site. All measurements marked with these factors in the NWIS dataset were removed. Some

2. Continuous monitoring wells have been used mostly since the early 1990s and use floats or other sensors to measure the altitude of the water table every 15-60 minutes (McGuire 2012)

measurements were recorded as negative values while the majority of measurements were entered as positive values. The negative values were changed to positive values to ensure continuity in the data and prevent averages from being calculated incorrectly.

The following steps were then used to estimate the yearly change in depth to groundwater at the county level. First, depth to water measurements at each site (well) were averaged over each winter period—October through May. Since each winter period includes two calendar year dates, for simplicity in the calculations, each winter period is identified by the calendar year beginning in October. For example, the winter period $t$ =2000 is the period October 2000 through May 2001. Next, the differences between subsequent winter period averages were calculated for each site.

To clarify these first two steps, let $m$ be the number of observations at well $j$ in period $t$ and let $n$ be the number of observations at well $j$ in period $t-1$. Then the change in depth to water at well $j$ in year $t$ is $delta_{j,t}$ shown below. Since the dataset begins with the calendar date October 1st, 1980 and ends in the calendar date May 31st, 2010 we have,

$$t \in \{1981, \dots, 2010\}$$

$$delta_{j,t} = \frac{1}{m} \sum_{i=1}^{m} depth_{i,j,t} - \frac{1}{n} \sum_{i=1}^{n} depth_{i,j,t-1}$$

Now, if there were no measurements in period $t-1$ or $t$ at well $t$ —hence, $m$ or $n$ is zero—then no delta is computed for year $j$ at that well site. Finally, to aggregate the well-level changes up to the county level, all $delta_{j,t}$ within a county $k$ in period $t$ were averaged. Explicitly, if there were $N_{k,t}$ many $delta_{j,t}$ in county $k$ in period $t$, then for all $j \in \{1,..,N_{k,t}\}$ the following county delta was computed.

$$county\,delta_{k,t} = \frac{1}{N_{k,t}} \sum_{j=1}^{N_{k,t}} delta_{k,t,j}$$

*2.3 Comparison with Existing Estimations*

This method has three important benefits over existing methods in the literature. These benefits all come at the expense of accuracy, however, the econometric model results in Chapter 3 demonstrate their validity. First, this procedure is easier and more quickly implemented than Theissen polygons[3] or raster interpolation methods[4], allowing for computation of yearly, county-level changes in depth to water over the 29-year observation period. Second, it eliminates inter-season variation caused by different wells being measured within a county from season to season. Third, the problem of well selection is easier, consistent and generalizable to other time periods and regions with similar USGS well data.

Theissen polygons and GIS raster files are the two most common interpolation methods used in the hydrology literature for large-scale estimates of water in storage and water table altitude or depth. These methods, both usually done in GIS, provide estimates for changes in depth to water by interpolating water table surface contours in each time period and finding the average change at the state, county, or regional level by differencing one contour from the other and dividing by the area of interest. These are presumed in this study to be the best methods available for finding county level changes

3. See Thiessen (1911)
4. See McGuire (2012)

in depth to water. However, they require extensive knowledge of the respective GIS methods, the wells measured, and the aquifer in general in order to select the correct well observations and run the interpolations accurately. Thus, they are usually  only done every 5 to 15 years since 1980 and every few decades before 1980 which is not frequent enough to be of maximum benefit to economic models that have yearly data for their other variables.

The depth to water can vary significantly within a county. Preventing the introduction of variability from this in the estimates[5] caused solely by the location and type of wells measured will be crucial to getting statistically significant results in econometric model estimation later on. This problem would arise, for example, if the more simple method were used of computing each average, county-level winter period depth using all of the well measurements (or all monitoring well sites) available. Matching sites between winters first, then differencing their averages, and finally averaging the differences across the county eliminates that problem.

A central problem for researchers using depth to water estimates in their statistical models—especially those outside the field of hydrology—is of coming up with or applying correct criteria for selecting well measurements to use. It is generally accepted to only use measurements taken during non-summer or growing season months such as October through May or November through April with the most extreme criteria found in the literature being December and January only. Well measurement locations are selected for many reasons such as not being close to a river or not close to municipal and other

5. Referred to henceforth as a “delta”, that is, the estimated average change in depth to water from one winter period to the next in a county

wells that pump during the winter. The well selection criteria in the new method presented above takes all these factors into account in a generalizable way without the need for one to come up with ad-hoc well selection rules for each individual study.[6]

## 3. Econometric Estimation

### *3.1 Variables and Model Review*

Three econometric approaches to panel data estimation are herein considered and discussed. The dependent variable in all of the models is the change in depth to water as calculated in the previous section. A negative value is a drop in the water table elevation and vice versa. The independent variables used are irrigation—"irr" measured as the percentage of the area of a county that is irrigated land, precipitation—"rain" measured as the inches of precipitation during the calendar year, temperature—"temp" measured as the number of degree days in a county over 62° F during the calendar year, and recharge—"rchg" estimated as the change in depth to groundwater attributed to recharge. All models were estimated in R using the econometric, panel linear models package "plm" (Croissant 2008).

The motivation for using panel data models comes from their ability to estimate and/or control for omitted variables in the model that we believe to exist but do not have data for. These omitted variables are referred to as "unobserved effects" and are discussed in Chapter 10 of Wooldridge (2002). There is intuitive reason to believe that there are unobserved effects in each county which are time-invariant or at least change very little

6. See Steward and Allen (2016, 38-39) for a summary of intra-season variation issues as well as a good discussion on well selection problems for their non-linear trend model

over the thirty year observation period. These would be factors like; typical crops grown or repeated crop rotations, lateral movement of groundwater from neighboring counties, types of soils and their specific yields under the county, population density[7], and proximity to surface water.

The three panel models used are Pooled OLS (POLS), Random Effects (RE), and Fixed Effects (FE). The POLS model uses the standard OLS estimator for the coefficients and puts the unobserved effect for each county in the error term. This county error term is corrected for by using a serial correlation robust variance estimator. The FE model puts the unobserved effect into the model as a county intercept. An OLS estimator is then used on a transformed version of the model and there is no serial correlation inherent in the model which allows it to be more robust to correlation between the unobserved effects and independent variables; a problem assumed away with the POLS and RE models. However, the FE model sacrifices the ability to put time constant independent variables in the model. The RE model is also estimated via OLS, but on a quasi-transformed model. Like POLS, serial correlation needs to be accounted for and it has more restrictive assumptions on the relationship between the unobserved effects and independent variables.

The POLS model is defined as,

$$y_{i,t} = x'_{i,t}\beta + v_{i,t}$$

where $v_{i,t} = c_i + u_{i,t}$ is the composite error term. Here and in all three model types, $c_i$ is the unobserved effect, the difference being in how it is treated. In POLS it is treated as a random error term that is time-invariant and the same across counties. Having $c_i$ in the

7. Assuming it is mostly constant relative to other counties' population density

composite error term leads to serial correlation which is corrected using a well-known, sandwich-type variance estimator robust to heteroskedasticity and serial correlation. This variance estimator is given in Wooldridge (2002, p.152, eq.7.26). The POLS estimator of the parameters $\boldsymbol{\beta}$ is identical to that of standard, multivariate OLS. A key assumption of both POLS and RE is that the conditional expectation meets the condition, $E[c_i; x_i]=0$ . This demands the independent variables to be uncorrelated with the unobserved effects of the error term.

The FE model is then defined as,

$$y_{i,t}=x'_{i,t}\beta+c_i+u_{i,t}$$

Despite the equation for the model looking similar to POLS, the way $c_i$ is defined is much different. In the FE model, $c_i$ is an intercept and not part of the error term. It can also be thought of as a dummy variable for each county or as a county-specific intercept. In the R "plm" package, extracting the fixed effects from the FE model yields the exact same values as obtained from the coefficients of the county dummy variables from a standard OLS model. Estimating a FE model with a county-specific intercept that is constant in time is known in the literature[8] and "plm" package as the "within" estimator. One can also include county fixed, time intercepts. Doing so, however, adds complexity in model interpretation with no intuitive reason for believing these effects to exist in this study. That is, the effects of irrigation and climate on depth to groundwater should not be effected by the time periods in which they were measured across all counties. From here on, the FE model, is assumed to be of the "within" type.

8. See Mundlak (1978) for further explanation and comparisons of these and related models not used here

The FE estimator then, can also be defined as the OLS estimator on the time-demeaned model,

$$y_{i,t} - \bar{y}_i = (x'_{i,t} - \bar{x}'_i)\beta + c_i - \bar{c}_i + u_{i,t}$$

Since $c_i$ is time invariant, $c_i - \bar{c}_i = 0$ which allows use of the OLS estimator. Note also an important drawback of time-demeaning; if any of the independent variables are time-invariant (like recharge) they will be removed from the model and their coefficients not estimated. For the purposes of this study, that drawback only applies to one independent variable which is recharge—an estimated average of county-level recharge taken during the 1980-2009 time period.

An important benefit of the FE model over POLS and RE is being able to relax the assumption, $E[c_i; x_i] = 0$ which says that the independent variables are uncorrelated with unobserved effects. Thus, in the FE model, the partial effects of the independent variables are estimated consistently given any relationship they have with the unobserved effects[9].

The RE estimator can be thought of, mathematically, as a weighted combination of the POLS and FE estimators where the RE estimator is an OLS estimator of the quasi-demeaned model,

$$y_{i,t} - \lambda \bar{y}_i = (x'_{i,t} - \lambda \bar{x}'_i)\beta + v_{i,t} - \lambda \bar{v}_i$$

and $\lambda$ varies from 0 to 1. Note that if $\lambda = 1$ then RE is the POLS estimator on the FE model and if $\lambda = 0$ then it is simply the POLS model with no demeaning. The

9. See Wooldridge (2002, 256-269) for complete discussion of assumptions required for POLS, RE, and FE models

Swamy and Arora (1972) method can also be thought of as a way to estimate $\lambda$ since $\lambda$ is a function of the variances of the two components of the error term. The POLS model is similar in definition to the RE model, but different in estimation of the parameters and variances. The POLS accounts for the unobserved effect in estimating the variance of the parameters, but estimates the parameters as if all the unobserved effects of each county are identical, hence, all counties are “pooled” together.

The RE model is defined as,

$$y_{i,t}=x'_{i,t}\beta+v_{i,t}$$

where $v_{i,t}=c_i+u_{i,t}$ is the composite error term like POLS. The differences being that the coefficients are estimated using estimates of each component of the composite error term rather than the OLS estimator and, likewise, the variances are estimated using sandwich-type estimators that estimate both error term components rather than just using the residuals as in the POLS variance estimator. There are multiple variance estimators known. The one used in this analysis is that of Swamy and Arora (1972), which is the default in the “plm” R package. The Swamy and Arora (1972) method can also be thought of as a way to estimate $\lambda$ since it is a function of the variances of the two components of the error term.[10]

The main drawback of employing all of these panel data models in this scenario is that they assume independence between the unobserved effects which may not hold. Lateral movement of water caused by natural factors or pumping implies that the water table under one county will be affected by its neighbors. However, the assumption of independence between counties only needs to hold between each time period for the

10. Explicitly, $\hat{\lambda}=f\left(\widehat{\sigma}_c,\widehat{\sigma}_u\right)$

model assumptions to be satisfied. Given the slow lateral movement of water it is plausible to assume that there is minimal interdependence between counties. However, cones of depression around wells close to county boundaries may cause unseen drawdowns in neighboring counties which could potentially violate the statistical independence of the counties in the regression. A further treatment of this system would be to use a covariance estimator designed to take spatial dependence into account. This violation aside, the major assumptions of the models above are plausible and, thus, the inferences below may be confidently compared to other studies in the literature.

### *3.2 Model Structure and Results*

Given the variables and models mentioned above, the most straightforward implementation would be a setup like,

$$delta_{i,t} = \left( rchg_i \,, irr_{i,t} \,, rain_{i,t} \,, temp_{i,t} \right) \beta + v_{i,t}$$

for POLS and RE while we would leave recharge out of the FE model because it is constant over time and would be eliminated during the FE time-demeaning transformation. Models of this type are called “plain” in the regression tables. They are included for comparison reason only.

The reason this model structure is not preferred is because of the understanding that rainfall and temperature that vary over the growing season do not directly affect variation in the delta variable. That is, rainfall in 2002 does not directly recharge the aquifer and cause a rise in the water table in the winter of 2003. This is because it takes many years for water at the surface to percolate down through the unsaturated zone to the water table (Scanlon et al. 2010, p. 12).[11] Soil moisture at the surface is also affected by pumping rates—due to irrigation return flow—and temperature—due to evapotransporation. Since soil moisture affects pumping rates via the producer's irrigation decisions to maximize crop yield, but does not affect the deltas via recharge—percolation to the water table—it is reasonable that if there is no irrigation in a county that the delta equals recharge. That is, if irrigation equals zero then delta equals recharge. The only simple linear panel models that satisfy these two conditions are,

$$delta_{i,t} = \left(rchg_i\ , irr_{i,t}\ , (irr * rain)_{i,t}\ , (irr * temp)_{i,t}\right) \beta + v_{i,t}$$

and

$$delta_{i,t} = \left(rchg_i\ , (irr * rain)_{i,t}\ , (irr * temp)_{i,t}\right) \beta + v_{i,t}$$

for POLS and RE. These are called "interact1" and "interact2" respectively. For the FE model the "rchg" variable is excluded, the county intercept $c_i$ is added, and the component error term $v_{i,t}$ is replaced by the idiosyncratic error term $u_{i,t}$.

Looking at the estimated pumping rates (Table 2) and recharge variable coefficient estimates (Tables 3, 4, and 5) give confirmation that models of the above, interacted form, are correct over the "plain" models. First, however, note that the "plain" models appear correct in the estimates of rain, temp, and irr coefficients. The irr variable should be negative since increasing the portion of a county covered in irrigated land should increase total pumping, hence, decreasing delta—lowering the altitude of the water table. The rain variable should be positive since increasing rainfall leads to a decrease in pumping rates, hence, an increase in delta—raising the altitude of the water

11. It takes at least 10 years for most counties and over 100 years in some areas of the Southern High Plains

table. The temp variable should be negative since increasing temperature increases evapotransporation which leads to an increase in pumping rates to maintain soil moisture, hence, decreasing delta. In Tables 3 through 5, the rain, temp, and irr coefficients are all significant over the 95% level in the plain models and have the correct signs.

However, including the rchg variable, it appears that the "plain" models start to lose validity against the interacted models. Since the rchg variable is the increase in water table elevation from recharge—not an increase in water in storage—it is expected to have a coefficient of 1. That is, recharge of one foot leads to an increase in delta of one foot. This variable is included mostly for testing the validity of the models and, as is shown in Tables 3 through 5, both interacted models in POLS and RE have a significant rchg coefficient close to 1, while the plain models are either not significant or well outside a neighborhood about 1. That is, the rchg coefficient in the plain models would all fail a hypothesis test with a null that the coefficient equals 1, even at the 90% level.

### *3.3 Inferences and Conclusions*

An additional advantage of the interacted model structure is that a linear pumping function is obtained using variables that have plentiful data already, without the need to estimate pumping at the county level over the last thirty years. This pumping rate would be in feet per year. For example, it can be written explicitly from the "interact2" model as,

$$\widehat{pumping} = \left( \widehat{\beta_{irr*rain}} * 1 * \bar{rain} + \widehat{\beta_{irr*temp}} * 1 * \bar{temp} \right) * specific\ yield$$

where the value of 1 is substituted in for the irrigation variable. The irrigation variable is defined as the percent of a county irrigated so a value of 1 corresponds to 100 percent of the county being irrigated. Thus, the estimated pumping rate is equal to the partial effect of irrigation on delta when a county is fully irrigated multiplied by the specific yield to get the acre-feet per acre application rate.

Specific yield ranges from 0.02 to 0.27 for almost all of the aquifer with 0.15 being the average (McGuire 2012a). The rest of the independent variable means are in Table 1 of the Appendix. Only the average pumping rate across all counties is examined here, however, one could look at the estimated pumping rates for specific counties by using county-level averages. Also, one could utilize the county intercept (estimated fixed effect) if the FE model is used.

Table 2 in the appendix presents the partial effect of irrigation on delta "dy_hat/dirr" given rain and temp means, estimated delta given no irrigation "y_hat | irr = 0", estimated delta given all of a county being irrigated "y_hat | irr = 1", and estimated pumping rates given a specific yield of 0.15 "pumping_hat". These pumping rates appear to be low and suggest the model estimates are somewhat biased. Estimated pumping rates from other sources range from 10 inches (0.8 feet) over the growing season in Nebraska to 16 inches (1.3 feet) in Texas. All of the estimated pumping rates are roughly half of the low average pumping rate in Nebraska of 0.8 feet. The POLS and RE pumping rates come out to just above 0.3 feet per year while the FE pumping rates come out to about 0.18 feet per year. The POLS and RE results are still reasonably close enough in proximity to the previous estimates since a high specific yield of 0.3, which occurs in some counties, would give a low but near estimate of 0.8 feet per year. The pumping rate estimates can still be used to justify using either of the interacted models over the "plain" models. Note that all three of the "plain" models give lower pumping estimates than their respective interacted model counterparts. The pumping estimates also suggest that the RE and POLS specifications, which have the unobserved effect in the error term, are more suitable since they are both quite closer to the correct pumping rate range of 0.8-1.3 than the pumping rates from the FE models.

Estimating a dependent variable in a novel way then using it to run econometric regressions on a hydrological system is bound to be somewhat dangerous. We are both incurring measurement error on the left-hand side by sacrificing accuracy for increased observation frequency while in the econometric estimation we are trying to use outside estimates of recharge and pumping rates to determine the validity of the model structure; all while using the same models to determine validity of the left-hand side variable estimates. The recharge variable and pumping rate comparisons don't fully justify the models implemented here, but it is apparent that they point to a move in the right direction from the standard "plain" model structure to the interacted model structure. Additionally, the R-squared values and estimated pumping rates suggest that an error-component model like POLS or RE is better specified than the FE models with only idiosyncratic error terms.

Another feasible and telling comparison can be made by estimating the average, aquifer-wide water-level change from pre-development[12] to 2011. This value can be

12. Pre-development is generally accepted to mean the period before 1950. Hence, the comparison looks at the change in average water-level from 1950-2011

obtained from each econometric model above by calculating $\hat{delta}$ using aquifer-wide means of each independent variable. This yearly, aquifer-wide $\hat{delta}$ value is then multiplied by 61 to get the estimated, cumulative change from 1950 to 2011. This value is then compared to USGS estimates in McGuire (2011 and 2012b). McGuire's average water-level change estimates come from differencing an interpolated, water-level contour from 2011 with one from the pre-development period. This is an important comparison because it uses all of the data and models discussed in this study to aggregate up to a single, aquifer-wide value and the methods and data are wholly different than McGuire's. The econometric models here are estimated using only 1980 through 2010 data whereas McGuire's (2011; 2012b) estimates use pre-development and 2011 data.

McGuire uses two different interpolation methods in the 2011 and 2012 reports and arrives at the total, area-weighted water-level changes of -14.2 and -14.4 feet respectively[13]. These values are assumed here to be the best estimates for pre-development to 2011 water-level changes in the HPA so the closer the estimates below are to these values the better that model is thought to perform. The estimated water-level change for each model is presented in the last column of Table 2 of the Appendix. As in the estimated pumping rate comparisons, the POLS and RE interacted models perform best with estimated changes of -15.75 and -15.73 feet respectively while the best FE interacted model gives -12.32 feet. Also, as with the pumping rate comparisons, the "plain" models perform the worst within each model type group. The plain models for POLS, RE, and FE come out to -12.24, -9.92, and -254.34 feet compared to the best interacted values above. Thus, the pumping rate comparisons and pre-development to

13. See McGuire (2012b, p.8 )for the complete table including state level changes

2011 water-level comparisons both indicate that the interacted setup is better than the plain setup and that the component error structure in POLS and RE performs better than the standard error structure and county intercept of the FE model. Moreover, the POLS and RE interacted models give total water-level change estimates that are very close to McGuire's estimates so these are a good starting point for future application of panel regression modeling of the aquifer.

Further investigation of the aquifer-wide system in this manner shows promise since this is only an initial attempt and did not nearly cover the gamut of econometric panel models which can take into account additional possibilities like variable coefficients and covariance matrices with a spatial inter-dependence structure that may be more applicable. Both of these would be feasible to estimate since the panel data set generated in this study has over 3000 observations which can accommodate a large number of independent variables and is not subject to common asymptotic problems that arise from many individuals and few time periods because T>>2 with the yearly water-level change estimates (delta variable). Other interesting possibilities would be to use these or similar estimates with long-run weather forecasts and irrigated acreage projections to predict aquifer depletion rates at the county, region or aquifer-wide level.

**References**

Croissant, Yves, Giovanni Millo, (2008), “Panel Data Econometrics in R: The plm Package”, Journal of Statistical Software, 27(2)

Gutentag, E.D., F.J. Heimes, N.C. Krothe, R.R. Luckey, and J.B. Weeks (1984), “Geohydrology of the High Plains aquifer in parts of Colorado, Kansas, Nebraska, New Mexico, Oklahoma, South Dakota, Texas, and Wyoming”, U.S. Geological Survey Professional Paper 1400(B):63

Heimes, F.J., and R.R. Luckey (1982), “Method for estimating irrigation requirements from ground water in the High Plains in parts of Colorado, Kansas, Nebraska, New Mexico, Oklahoma, South Dakota, Texas, and Wyoming”, U.S. Geological Survey Water-Resources Investigations Report, 82(40):64

High Plains Study Council (1982), “Summary of Results of the Ogallala Aquifer Regional Study, with Recommendations to the Secretary of Commerce and Congress”, Economic Development Administration, Washington D.C.

Kromm, David E. and Stephen E. White (1986), “Public Preferences for Recommendations Made by the High Plains-Ogallala Aquifer Study”, Social Science Quarterly, 67(4):841

McGuire, V.L. (2011), “Water-Level Changes in the High Plains Aquifer, Predevelopment to 2009, 2007–08, and 2008–09, and Change in Water in Storage, Predevelopment to 2009”, Scientific Investigations Report, 2011(5089)

McGuire, V.L., K.D. Lund, and B.K. Densmore, (2012a), “Saturated thickness and water in storage in the High Plains aquifer, 2009, and water-level changes and changes in water in storage in the High Plains aquifer, 1980 to 1995, 1995 to 2000, 2000 to 2005, and 2005 to 2009”, U.S. Geological Survey Scientific Investigations Report 2012(5177):28

McGuire, V.L. (2012b), “Water-Level and Storage Changes in the High Plains Aquifer, Predevelopment to 2011 and 2009–11”, Scientific Investigations Report, 2012(5291)

Mundlak, Yair (1978), “On the Pooling of Time Series and Cross Section Data”, Econometrica, 46(1):69-85

Qi, S. L., A. Konduris, D. W. Litke, and J. Dupree (2002), “Classification of irrigated land using satellite imagery, the High Plains Aquifer, nominal data 1992”, U.S. Geol. Surv. Water Resour. Invest. Rep., 02–4236

Scanlon, B.R., Ian Jolly, Marios Sophocleous, and Lu Zhang (2007), "Global Impacts of Conversions from Natural to Agricultural Ecosystems on Water Resources: Quantity versus Quality", Water Resources Research, 43(W03437)

Scanlon, B.R., R.C. Reedy, and J.B. Gates, (2010), "Effects of irrigated agroecosystems: 1. Quantity of soil water and groundwater in the Southern High Plains, Texas", Water Resources Research, 43(W09537)

Steward, David R., Paul J. Bruss, Xiaoying Yang, Scott A. Staggenborg, Stephen M. Welch, and Michael D. Apley, (2013), "Tapping unsustainable groundwater stores for agricultural production in the High Plains Aquifer of Kansas, projections to 2110, PNAS 110(37):E3477-E3486

Steward, David R., Andrew J. Allen, (2016), "Peak groundwater depletion in the High Plains Aquifer, projections from 1930 to 2110", Agricultural Water Management, 170:36-48

Swamy P.A.V.B. and S.S. Arora (1972). "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models." Econometrica, 40:261

Thiessen, A.H., (1911), "Precipitation averages for large areas", Monthly Weather Review, 39:1082-1084

Wooldridge, Jeffrey M., (2002), "Econometric Estimation of Cross Section and Panel Data", The MIT Press, Cambridge, Massachusetts

**Appendix**

*Figure 1*

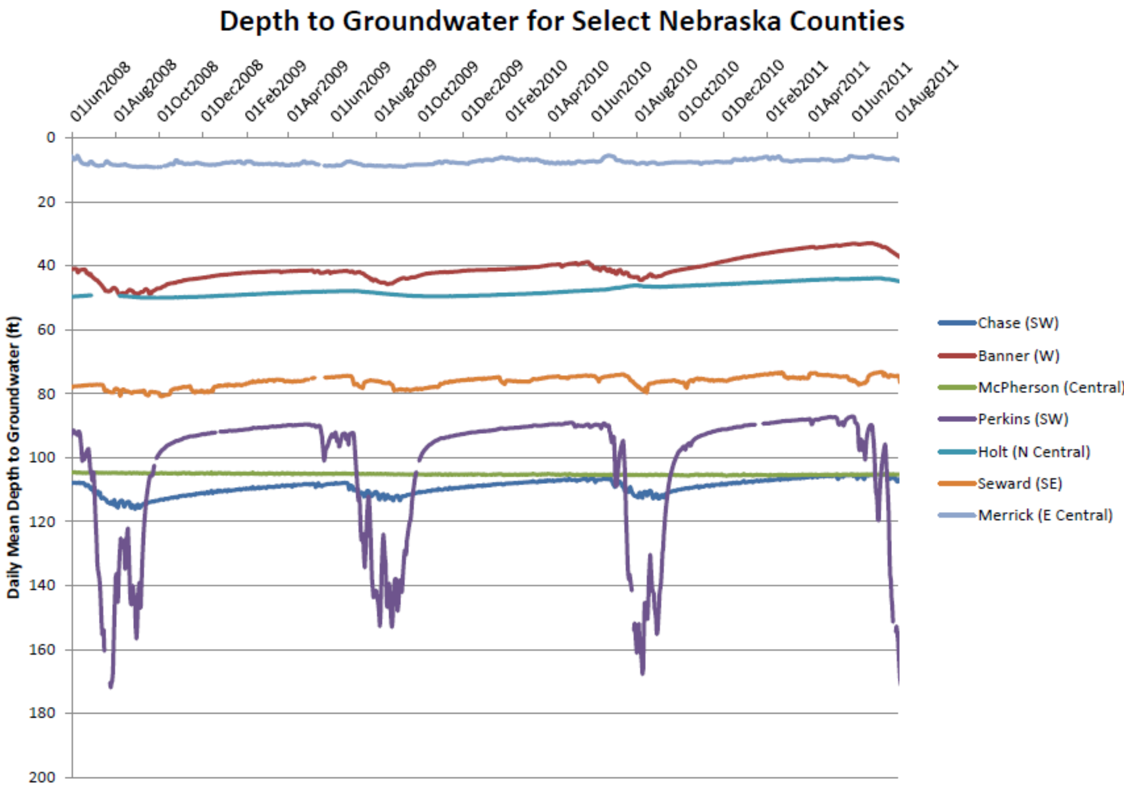


*Table 1*

| | mean |
|---|---|
| constant | 1 |
| rchg | 0.150024 |
| irr | 0.1361112 |
| rain | 23.551 |
| temp | 53.24 |

*Table 2*

(all values in feet)
*no means or coefficients were used to calculate these values

| | dy_hat/dirr | y_hat \| irr=0 | y_hat \| irr=1 | pumping_hat | predev to 2011 |
|---|---|---|---|---|---|
| ols_plain | -1.18 | -0.04 | -1.22 | 0.18 | -12.24 |
| ols_int1 | -2.07 | 0.02 | -2.05 | 0.31 | -15.75 |
| ols_int2 | -2.07 | 0.02 | -2.04 | 0.31 | -15.75 |
| rand_plain | -1.49 | 0.04 | -1.45 | 0.22 | -9.92 |
| rand_int1 | -2.16 | 0.04 | -2.12 | 0.32 | -15.73 |
| rand_int2 | -2.15 | 0.03 | -2.12 | 0.32 | -15.77 |
| fixed_plain | -0.96 | -4.04 | -5.00 | 0.14 | -254.34 |
| fixed_int1 | -1.48 | 0.00* | -1.48 | 0.22 | -12.32 |
| fixed_int2 | -1.19 | 0.00* | -1.19 | 0.18 | -9.87 |

*Table 3*

| | OLS Plain | OLS Interact 1 | OLS Interact 2 |
|---|---|---|---|
| Constant | $2.629^{***}$ | $-0.135^{***}$ | $-0.134^{***}$ |
| | (1.953, 3.305) | (−0.219, −0.050) | (−0.221, −0.047) |
| rchg | $0.416^{*}$ | $1.052^{***}$ | $1.049^{***}$ |
| | (−0.043, 0.876) | (0.642, 1.462) | (0.605, 1.492) |
| irr | $-1.182^{***}$ | | 0.087 |
| | (−1.466, −0.898) | | (−3.800, 3.973) |
| rain | $0.076^{***}$ | | |
| | (0.070, 0.082) | | |
| temp | $-0.085^{***}$ | | |
| | (−0.097, −0.073) | | |
| irr:rain | | $0.423^{***}$ | $0.423^{***}$ |
| | | (0.397, 0.449) | (0.393, 0.453) |
| irr:temp | | $-0.226^{***}$ | $-0.227^{***}$ |
| | | (−0.240, −0.212) | (−0.295, −0.160) |
| Observations | 3,505 | 3,505 | 3,505 |
| $R^2$ | 0.215 | 0.258 | 0.258 |
| Adjusted $R^2$ | 0.214 | 0.257 | 0.257 |

*Table 4*

| | Fixed Plain | Fixed Interact 1 | Fixed Interact 2 |
|---|---|---|---|
| irr | $-0.965^{*}$ | | $7.215^{*}$ |
| | $(-2.043, 0.114)$ | | $(-0.035, 14.465)$ |
| rain | $0.147^{***}$ | | |
| | $(0.138, 0.155)$ | | |
| temp | $-0.141^{***}$ | | |
| | $(-0.170, -0.112)$ | | |
| irr:rain | | $0.514^{***}$ | $0.500^{***}$ |
| | | $(0.480, 0.547)$ | $(0.464, 0.536)$ |
| irr:temp | | $-0.255^{***}$ | $-0.379^{***}$ |
| | | $(-0.279, -0.231)$ | $(-0.506, -0.252)$ |
| Observations | 3,505 | 3,505 | 3,505 |
| $R^2$ | 0.313 | 0.221 | 0.222 |
| Adjusted $R^2$ | 0.282 | 0.187 | 0.187 |

*Table 5*

| | Random Plain | Random Interact 1 | Random Interact 2 |
|---|---|---|---|
| Constant | $3.533^{***}$ | $-0.108^{*}$ | $-0.103$ |
| | $(2.471, 4.594)$ | $(-0.235, 0.019)$ | $(-0.232, 0.026)$ |
| rchg | $-0.398$ | $0.957^{***}$ | $0.914^{***}$ |
| | $(-1.242, 0.445)$ | $(0.322, 1.591)$ | $(0.240, 1.587)$ |
| irr | $-1.485^{***}$ | | $0.937$ |
| | $(-1.989, -0.981)$ | | $(-4.046, 5.921)$ |
| rain | $0.115^{***}$ | | |
| | $(0.108, 0.123)$ | | |
| temp | $-0.116^{***}$ | | |
| | $(-0.134, -0.097)$ | | |
| irr:rain | | $0.462^{***}$ | $0.459^{***}$ |
| | | $(0.433, 0.491)$ | $(0.426, 0.492)$ |
| irr:temp | | $-0.245^{***}$ | $-0.261^{***}$ |
| | | $(-0.261, -0.229)$ | $(-0.349, -0.173)$ |
| Observations | 3,505 | 3,505 | 3,505 |
| $R^2$ | 0.262 | 0.233 | 0.233 |
| Adjusted $R^2$ | 0.261 | 0.232 | 0.232 |