

A relação inversa entre o tamanho das propriedades agrícolas e a produtividade no Brasil: uma análise não paramétrica usando regressões *kernel*

The inverse relationship between farm size and productivity in Brazil: a non-parametric analysis from kernel regressions

Alexandre Amorim Souza Ferreira¹ , Alexandre Nunes Almeida¹ 

¹Programa de Pós-graduação em Economia Aplicada, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo (USP), Piracicaba (SP), Brasil. E-mails: alexandre_amorim2005@hotmail.com; alex.almeida@usp.br

Como citar: Ferreira, A. A. S., & Almeida, A. N. (2021). A relação inversa entre o tamanho das propriedades agrícolas e a produtividade no Brasil: uma análise não paramétrica usando regressões *kernel*. Revista de Economia e Sociologia Rural, 59(3), e224128. <https://doi.org/10.1590/1806-9479.2021.224128>

Resumo: O principal objetivo deste artigo é examinar a hipótese da relação inversa entre o tamanho das propriedades agrícolas e a produtividade no Brasil usando métodos de regressão *kernel* não paramétricos. Enquanto os estimadores paramétricos são considerados globais, as regressões *kernels* usam uma amostra de dados próximas a um ponto para ajustar a estimação, o que permite focar em peculiaridades locais da distribuição dos dados, permitindo que os dados "falem por si mesmos". Ambos os métodos foram aplicados e comparados a partir de dados agregados do Censo Agropecuário de 2006. Entre os principais resultados, observou-se que a hipótese da relação inversa entre a área e a produtividade não é corroborada na análise paramétrica; no entanto, quando os diferentes estimadores não paramétricos são comparados, para a maioria das regiões do país, a relação inversa é observada nas menores propriedades, mas não nas médias e grandes.

Palavras-chave: relação inversa, produtividade, tamanho da propriedade, regressão *kernel* não paramétrica.

Abstract: The main objective of this work is to examine the hypothesis of the inverse relationship between farm size and productivity in Brazil using non-parametric kernel methods. While parametric estimators are considered global, kernel regressions use a sample of data close to a point to adjust the estimation, which allows focusing on local peculiarities of the data distribution, allowing the data "to speak for themselves". Both methods are applied and compared using aggregated data from the 2006 Agricultural Census. Among the main results, it was observed that the hypothesis of the inverse relationship between area and productivity was not corroborated in the fully parametric analysis; however, when the different non-parametric estimators are compared, the inverse relationship hypothesis is observed for small farmers, but not for medium and large farmers in most regions of the country.

Keywords: inverse relationship, productivity, farm size, kernel regression.

1. Introdução

A relação que existe entre o tamanho da propriedade e sua produtividade agrícola é um tema antigo e amplamente discutido na literatura econômica (Eastwood et al., 2010). Em resumo, se ela for direta, um aumento na área cultivada irá promover um aumento na produtividade em maior magnitude que o aumento observado da área; e, quando inversa, um aumento na área terá, como consequência, um aumento de produtividade menor (ou ainda até uma redução de produtividade) (Benjamin, 1995). Em outras palavras, quando a



relação observada é inversa, implica que menores fazendas têm produção por medida de área maior do que propriedades maiores (Lamb, 2003).

O conhecimento empírico tem sugerido que esta relação área versus produtividade seja negativa para países com menor nível de desenvolvimento (Feder, 1985; Chen et al., 2011; Ali & Deininger, 2015). A relação inversa foi observada pela primeira vez na Rússia por Chayanov (1926), no entanto, as primeiras citações sobre o assunto são creditadas a Sen (1962), Bardhan (1973), e Yotopoulos & Lau (1973) e Carter (1984) para a Índia. Na década de 1970, autores como Sen (1975), Berry & Cline (1979) escreveram livros¹ que se tornaram referências importantes, enquanto a partir da década de 1980 observou-se um grande aumento de publicações e de estudos empíricos sobre o tema², que tem sido objeto de debates prolongados e acalorados (World Bank, 2008; Julien et al., 2019).

Várias explicações para tentar esclarecer completamente a hipótese da relação inversa têm sido testadas ao longo dos anos, mas nenhuma delas ainda foi suficiente para alcançar resultados conclusivos (Verschelde et al., 2013; Ali & Deininger, 2015). Primeiro, sugere-se que as imperfeições que existem no mercado de trabalho (Feder, 1985; Heltberg, 1998; Ali & Deininger, 2015), de crédito (Binswanger & Rosenzweig, 1986; Heltberg, 1998; Assunção & Ghatak, 2003), de seguros (Binswanger et al., 1995) e de terras (Heltberg, 1998) poderiam ser apontadas como importantes razões para se observar a relação inversa. Assim, as falhas, ou mesmo ainda a ausência de algum destes mercados, geram distorções na alocação de recursos nas propriedades agrícolas e implicam ineficiências na atividade produtiva, resultado da subutilização da terra (Verschelde et al., 2013). Mais especificamente, em grandes fazendas, trabalhadores quando contratados não compartilham os mesmos riscos como se fossem donos da área, são menos motivados e se esforçam menos, gerando um alto ônus de supervisão dessa mão de obra aos fazendeiros (Feder, 1985; Heltberg, 1998; Henderson, 2015). Por outro lado, instituições financeiras, devido à assimetria de informações e ao risco da agricultura, racionam o crédito (Carter & Wiebe, 1990) e requerem garantia para a realização de empréstimos, dificultando o acesso aos produtores que possuem áreas menores (Binswanger & Rosenzweig, 1986). Binswanger et al. (1995) destacam que há ainda a falta de um mercado de seguros para risco de preços nos países em desenvolvimento, enquanto sua priorização é direcionada para os maiores produtores. Por fim, Heltberg (1998) menciona os problemas de falhas no mercado de terras de aluguel e venda. Segundo Heltberg (1998), o arrendatário tende a usar menos insumos e subestimar a renda se o contrato é compartilhado. Há ainda o risco de quebra de safra e o não pagamento do contrato, se é pago pós-colheita. Por outro lado, se pago antes, pode ocorrer falta de liquidez e impedir o acesso a crédito, reduzindo o potencial de investimento na produção. Já as barreiras impostas pelos financiamentos, como juros altos e risco moral, restringem o acesso para aquisição de terras por parte dos pequenos e médios produtores (Heltberg, 1998).

Outra explicação bastante comum na literatura é a relação inversa causada por vieses de variáveis omitidas, por exemplo, a qualidade do solo (Berry & Cline, 1979; Bhalla, 1988; Benjamin, 1995; Carletto et al., 2013). Diferenças na qualidade do solo geram diferenças na produção. Entretanto, o controle para esta variável não foi capaz de explicar a relação inversa para o Brasil (Berry & Cline, 1979) e África (Barrett et al., 2010; Carletto et al., 2013; Ali & Deininger, 2015), enquanto ela desapareceu na China (Chen et al., 2011).

Algumas características dos proprietários agrícolas, que são inatas a estes (ex., motivação), também são consideradas como variáveis omitidas (Assunção & Ghatak, 2003; Assunção & Braido, 2007). Assunção & Braido (2007) investigaram a influência desta variável na Índia, entre outras informações demográficas, como sexo do chefe, número de componentes e razão de dependência familiar; no entanto, não foram suficientes para explicar a relação inversa. Além disso, outras evidências para outros países também sugerem que diferenças nas habilidades gerenciais inatas observadas entre os produtores também não explicam totalmente a relação inversa (Lipton, 2009; Verschelde et al., 2013).

¹ Sen (1975) escreveu o livro *Employment, Technology and Development*, enquanto Berry & Cline (1979) o *Agrarian structure and productivity in developing countries*.

² Uma revisão completa sobre o assunto está fora do escopo deste trabalho, mas pode ser encontrada em Lipton (2009) e Eastwood et al. (2010).

Questões metodológicas são a terceira explicação tradicional encontrada na literatura (Verschelde et al., 2013). Por exemplo, Carletto et al. (2013) investigaram se os erros de mensuração e arredondamentos das áreas autodeclaradas, ao invés de usar informações georreferenciadas, poderiam ser a causa da relação inversa para Uganda, cuja conclusão foi negativa. Lipton (2009) também afirma que as propriedades de tamanhos diversos não podem ser consideradas “réplicas” uma das outras, o que pode gerar viés nas estimações, visto que elas não são. É conhecido o fato que as propriedades maiores utilizam tecnologias diferentes que as menores e o uso de insumos varia com a escala da produção. Assim, modelos econométricos que não levem em consideração a verdadeira estrutura dos dados, e estão sujeitos a mesma especificação (parâmetros), podem falhar ao capturar erroneamente as não linearidades dos dados de uma amostra se estes forem significativamente heterogêneos (Verschelde et al., 2013).

Este trabalho tem como objetivo testar se existe uma relação inversa entre o tamanho das propriedades agrícolas brasileiras e a produtividade destas usando dados do Censo Agropecuário do ano de 2006 coletados pelo Instituto Brasileiro de Geografia e Estatística (IBGE). A novidade na nossa análise é adoção de métodos de regressão *kernel* não paramétrico. Uma das grandes vantagens da abordagem das regressões não paramétricas é levar em consideração as peculiaridades locais que os dados possam apresentar sem assumir nenhuma suposição *a priori* sobre a função a ser estimada, isto é, permitir que os dados falem por si mesmos (Eubank, 1999).

Testar se existe uma relação inversa entre o tamanho das propriedades agrícolas e a produtividade é de suma importância, porque colabora para que políticas de redistribuição de terras e projetos de desenvolvimento agrícola para melhorar a produção e a equidade sejam implementados de acordo (World Bank, 2008). No entanto, sabe-se também que reformas agrárias tendem a estar associadas a um distúrbio social significativo e difíceis de implementar (Eastwood et al., 2010). Como resultado, políticas agrícolas implementadas em várias economias em desenvolvimento nas últimas duas décadas que evitaram a redistribuição de terras, e procuraram corrigir falhas nos mercados e vincular pequenos produtores mais às modernas cadeias de valor agroalimentar, contribuíram mais positivamente para diminuir as desvantagens de produtividade existentes entre maiores e menores propriedades (Ali & Deininger, 2015; Rada & Fuglie, 2019).

Seguindo a liberalização dos mercados e do comércio nos anos 1980 e 1990, Deininger & Byerlee (2012), Helfand et al. (2015) e Rada & Fuglie (2019) enaltecem que o Brasil apresenta diversas políticas relacionadas às grandes propriedades, fomentadas pelo Ministério da Agricultura, e também às pequenas propriedades (propriedades familiares) pelo Ministério do Desenvolvimento Agrário.³ Esta dualidade de políticas, às vezes com distorções, faz do Brasil um interessante objeto para este estudo, que somada à importância de se entender a relação entre o tamanho e a produtividade das propriedades agrícolas, constitui-se como a principal motivação do presente trabalho. Ademais, a adoção da abordagem de regressão *kernel* não paramétrico será uma grande contribuição, visto que até o presente ainda não foram encontrados trabalhos na literatura brasileira que adotem este método para este tipo de análise.

Além desta introdução, este trabalho está apresentado da seguinte forma: a seção seguinte ilustra de forma sucinta a literatura relacionada, bem como apresenta as motivações para o uso da regressão *kernel* não paramétrica para atender aos objetivos propostos pelo trabalho. Na terceira seção, a estratégia empírica adotada é apresentada, enquanto na quarta são apresentados os dados utilizados na análise. Na quinta, os resultados são discutidos e, por fim, na sexta seção, as conclusões do artigo.

³ A Secretaria Especial de Agricultura Familiar e do Desenvolvimento Agrário da Casa Civil da Presidência da República (Sead) é atualmente responsável pelas competências anteriormente atribuídas ao Ministério do Desenvolvimento Agrário (MDA), extinto em 12 de maio de 2016 através da medida provisória nº 726 (Brasil, 2016).

2. Literatura relacionada e motivação para a regressão *kernel* não paramétrica

No Brasil, a quantidade de estudos que buscaram analisar os determinantes da evolução da produtividade agrícola no país ou em regiões específicas é extensa (Vicente et al., 2000; Mendes et al., 2009; Bragagnolo et al., 2010; Magalhães et al., 2011; Ávila et al., 2013; Felema et al., 2013; Gasques et al., 2012; Rada & Buccola, 2012; Rada & Valdes, 2012; Gasques et al., 2014; Bragagnolo & Barros, 2015; Ferreira et al., 2016; Carrer & Souza Filho, 2016; Soares & Spolador, 2019). No entanto, estes mesmos autores não correlacionam seus achados com a literatura da hipótese da relação inversa, sendo o tema ainda pouco explorado no país. Enquanto Berry & Cline (1979) concluíram que a produção por área agrícola diminui à medida que o tamanho da fazenda aumenta na região Nordeste, Helfand & Levine (2004) e Rada et al., (2019), utilizando dados dos censos agropecuários, para o Centro-Oeste e Brasil, respectivamente, encontraram uma relação entre a produtividade e o uso da terra em forma de U, indicando que propriedades médias seriam menos eficientes que as pequenas e grandes. Lázari & Magalhães (2016) procuraram correlacionar o crescimento da produtividade total dos fatores (PTF) conforme o tamanho dos estabelecimentos rurais na região Sudeste a partir dos dados dos censos agropecuários e de outras fontes. Eles mostraram que propriedades acima de 100 ha alcançaram um crescimento da PTF maior do que as propriedades menores que 100 ha, não corroborando, portanto, a relação inversa.

Assim, dado que a linearidade dos parâmetros é assumida pela maior parte destes autores supracitados, é nossa motivação investigar para o Brasil se existe uma relação inversa entre tamanho e produtividade agrícola quando métodos *kernel* não paramétricos são adotados ao estimar funções de produção.

A principal vantagem do método de regressão *kernel* não paramétrico é permitir estimar uma função sem predefinir a sua forma funcional (Henderson & Parmeter, 2015), ou seja, permitir aos dados falarem por si mesmos (Eubank, 1999). Henderson & Parmeter (2015) explicam que, enquanto os estimadores paramétricos são considerados globais (usam todos os pontos), os estimadores *kernels* não paramétricos usam uma amostra dos dados próxima a um ponto para ajustar a estimação, e a partir deste ajuste constrói-se um estimador global. A grande vantagem aqui é que isto permite focar nas peculiaridades locais relativas ao banco de dados, enquanto a função desconhecida (sem forma funcional preestabelecida) é estimada. Em resumo, o valor médio da variável dependente é calculado em valores específicos de cada covariável.

Verschelde et al. (2013) demonstram que a média condicional ($E[Y|X=x]$) é estimada pelas médias de regressões localizadas ($E[Y|X \text{ próximo de } x]=g(x)$). Assim sendo, o emprego da regressão *kernel* não paramétrica para analisar a relação entre o tamanho das propriedades agrícolas e a produtividade: 1) terá estimativas independentes de formas funcionais; 2) irá mostrar a forma sem restrições na estrutura de formação dos dados desta relação (não linearidades); 3) trará informações locais (se há diferenças na relação entre os tamanhos das propriedades) e globais (considerando todas as propriedades) ao mesmo tempo.

Contudo, como todos os métodos estatísticos, o método de regressão não paramétrico também apresenta desvantagens. Devido ao uso de uma amostra de dados próximas a um ponto para a estimação do modelo global, quanto maior a dimensão, maior a dificuldade para o estimador perceber a real estrutura dos dados. Este fenômeno estatístico é chamado de “maldição da dimensionalidade” (Li & Racine, 2007). Henderson & Parmeter (2015) definem este fenômeno como sendo a dificuldade em detectar a estrutura do processo de geração de dados em altas dimensões sem excessivas suposições paramétricas. Em uma dimensão é fácil perceber a estrutura dos dados, visto que eles estarão dispostos em uma reta, sendo necessário o modelo somente preencher o espaço entre um ponto e outro (sendo com uma constante ou uma linha) horizontalmente. Quando a dimensão aumenta para duas, faz-se necessário preencher o espaço entre um ponto e outro, tanto horizontalmente quanto verticalmente, visto que eles não estarão mais dispersos em uma linha, e sim em um eixo cartesiano. A Figura 1 busca ilustrar melhor esta questão em uma e duas dimensões.

de arroz por hectare como variáveis dependentes e a área cultivada como independente. A relação inversa foi constatada em ambos os países.

Verschelde et al. (2013) usou um banco de dados para Burundi, que continha variáveis contínuas e discretas para estimar a relação inversa. Foi estimada uma função de produção não paramétrica com o logaritmo do valor da produção por hectare como medida de produtividade, o logaritmo da área como principal variável explicativa, e outras variáveis relativas às características das famílias e das propriedades. A relação inversa foi observada e suportaram a hipótese de que a relação é inversa.

Ali & Deininger (2015), através de uma regressão *kernel* não paramétrica, analisaram a relação do valor da produção da colheita por hectare e dos lucros por hectare com o tamanho das parcelas das propriedades, todos em logaritmos, para Ruanda. Em ambos os casos, a relação inversa foi encontrada.

3. Estratégia Empírica

Em geral, a teoria econômica não fornece muita orientação sobre a escolha da forma funcional para fins de estimação de qualquer modelo econométrico (Yatchew, 2003; Henderson & Parmeter, 2015). De acordo com Blundell & Duncan (1998), existe também uma obsessão por linearidade por parte dos economistas empíricos, embora esta escolha seja justificada por ser de fácil interpretação em relação aos modelos de regressão não paramétricos (Henderson & Parmeter, 2015). No entanto, modelos não paramétricos ou semiparamétricos constituem um avanço importante, já que permitem verificar se a especificação da função é adequada de acordo com a natureza dos dados (Henderson & Parmeter, 2015).

Desta maneira, para ilustrar a diferença entre as abordagens paramétrica e não paramétrica ao testar a relação entre a produtividade e a área, considere, primeiramente, uma forma funcional simplificada totalmente paramétrica para cada propriedade agrícola i como:

$$\ln Y_i = \alpha + \beta \ln A_i + \gamma \ln X_i + u_i \quad (1)$$

onde α, β, γ são os parâmetros a serem estimados e u é o termo de erro. Segundo Lamb (2003), se β mede a elasticidade do produto Y ou lucros em relação à área A , quando $\beta < 1$, sugere que o produto total aumenta menos rapidamente do que a área cultivada, logo, a relação é considerada inversa. Mas, se Y for a medida de produtividade ou valor da produção por área cultivada, a relação é considerada inversa se $\beta < 0$ (Heltberg, 1998). O vetor X_i inclui outras características-chaves relacionadas com a produção agrícola, como gastos com capital e insumos, e mão de obra familiar e contratada.

Agora, considere o seguinte modelo totalmente não paramétrico como:

$$\ln Y_i = \varphi(\ln A_i, \ln X_i) + u_i \quad (2)$$

onde a Equação 2 segue a mesma especificação exibida na Equação 1, exceto que agora $\varphi(\cdot)$ uma função desconhecida de A_i, X_i . Para estimar a Equação 2, três estimadores não paramétricos são adotados no presente trabalho para checagem da robustez das estimativas, a saber, 1) o estimador constante-local de mínimos-quadrados (LCLS), 2) o estimador linear-local de mínimos-quadrados (LLLS), e 3) o estimador quadrático-local de mínimos quadrados (LQLS) (Henderson & Parmeter, 2015).

Todas as propriedades teóricas para cada um destes três estimadores supracitados estão apresentadas em Li & Racine (2007) e Henderson & Parmeter (2015), mas omitidas aqui no presente estudo. Henderson & Parmeter também fornecem códigos de software R que permitem aos usuários replicarem todos os exemplos do livro deles de 2015. Os códigos podem ser acessados em <http://www.the-smooth-operators.com/code>.

Cabe salientar que o ponto de partida destes três estimadores *kernel* adotados pelo presente trabalho partiu do estimador não paramétrico desenvolvido por Nadaraya (1964) e Watson (1964). A título de ilustração, como exemplo, considere a Equação 2 em sua forma univariada e sem logaritmo como:

$$Y_i = \varphi_i(A_i) + u_i. \quad (3)$$

Assim, o estimador Nadaraya-Watson, baseado na Equação 3, pode ser expresso por:

$$\hat{\varphi}_i(A_i) = \frac{\frac{1}{nh} \sum_{i=1}^n k\left(\frac{A_n - A_i}{h}\right) Y_i}{\frac{1}{nh} \sum_{i=1}^n k\left(\frac{A_n - A_i}{h}\right)} \quad (4)$$

onde h é largura da janela, n o número de observações e $k(\cdot)$ é uma função *kernel* que atende a todas as propriedades estatísticas de: simetria, integração para 1, e continuamente diferenciável (Henderson & Parmeter, 2015). As especificações das funções *kernel* mais utilizadas são: uniforme, gaussiana e epanechnikov (Li & Racine, 2007).

Por fim, para qualquer estimativa não paramétrica *kernel*, a escolha da largura da janela h é mais importante do que a escolha da função *kernel*, já que a largura da banda h regula o *trade-off* entre variância e viés nas estimativas (Li & Racine, 2007). Mais especificamente, a escolha ótima de h é feita através da minimização do erro assintótico integrado quadrático médio (AMISE). Menores larguras das janelas são desejáveis quanto maior o tamanho da amostra, pois maiores informações disponíveis permitem o estimador ser mais local (Li & Racine, 2007). Nestes casos, a técnica conhecida como validação cruzada de mínimos quadrados (LSCV-*Least Squares Cross Validation*) está entre os algoritmos mais populares para a seleção ótima da largura da janela h (Henderson & Parmeter, 2015), e que será adotada no presente estudo.

O objetivo da LSCV é minimizar a diferença entre o estimador da densidade e a própria densidade com relação à largura da janela sendo expresso, no caso univariado, por:

$$LSCV_{(h)} = \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n \bar{k}\left(\frac{A_i - A_j}{h}\right) - \frac{2}{n(n-1)h} \sum_{i=1}^n \sum_{j=1, j \neq i}^n k\left(\frac{A_i - A_j}{h}\right) \quad (5)$$

onde $k(\cdot)$ é função *kernel* (ex., gaussiana, epanechnikov, etc.), enquanto \bar{k} é chamada de convolução de *kernel*.⁵

4. Dados

Os dados utilizados foram coletados do Censo Agropecuário do ano de 2006, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e recuperados através do Sistema IBGE de Recuperação Automática de Dados (SIDRA) (Instituto Brasileiro de Geografia e Estatística, 2016). O censo foi realizado entre 2006 e abril de 2007 através da aplicação de questionários eletrônicos em cada um dos estabelecimentos agropecuários visitados pelo IBGE, de modo a coletar informações sobre as atividades agropecuárias desenvolvidas.

Por fins de confidencialidade, o IBGE não disponibiliza publicamente os dados por estabelecimento, mas os agrega por município e faixas de área. Cada município tem seus estabelecimentos agregados de acordo com a área, sendo 17 faixas de área definidas pelo

⁵ O capítulo 2 do livro de Henderson & Parmeter (2015) apresenta em detalhes o método LSCV entre outros métodos para a escolha ótima de h .

IBGE.⁶ As faixas de área que não têm estabelecimentos produtivos foram excluídas da análise. A análise econométrica explorou a variação entre elas. Assim sendo, o tamanho da amostra é potencialmente aumentado de 5.538 observações (relativo ao número de municípios coletados pelo IBGE) para 94.146 (número de municípios multiplicado pelo número de faixas de área de cada município).

As variáveis utilizadas na análise do presente trabalho são as mesmas para ambos os modelos paramétricos e não paramétricos, a fins de comparação entre eles. Em resumo, as variáveis são: uma medida de produtividade e insumos utilizados (área produtiva, capital e trabalho).

O valor da produção (VP em mil reais) por hectare é utilizado como medida da produtividade da área. O IBGE considera como sendo o valor da produção animal, vegetal e da agroindústria. A produção animal é subdividida em grande, médio e pequeno porte, além das aves. A vegetal em lavouras temporárias e permanentes, horticultura, floricultura, silvicultura e extração vegetal. Além disso, o valor da produção é dividido pela área produtiva das propriedades, mensurada em hectares, e transformados em logaritmos. As observações cujo valor da produção é 0 foram excluídas para evitar vieses resultantes de propriedades não produtivas que estão fora do escopo do trabalho.

De acordo com o objetivo proposto, a principal variável independente de interesse é o logaritmo da área produtiva das propriedades, a qual é mensurada em hectares. O IBGE disponibiliza a área total e desagrega em 15 utilizações diferentes para a terra como: 1) lavouras permanentes; 2) lavouras temporárias; 3) área plantada com forrageiras de corte; 4) área para cultivo de flores; 5) pastagens naturais; 6) pastagens plantadas; 7) pastagens plantadas degradadas; 8) pastagens em boas condições; 9) matas e/ou florestas naturais destinadas à preservação permanente ou reserva legal (excluídas as áreas de preservação permanente e as em sistemas agroflorestais; 10) matas e/ou florestas plantadas com essências florestais; 11) sistemas agroflorestais; 12) tanques e áreas de águas para exploração de aquicultura; 13) construções e benfeitorias; 14) terras degradadas; e 15) inaproveitáveis.

Foram excluídas observações que não eram consideradas produtivas (matas e/ou florestas destinadas à preservação permanente e reserva legal, matas e/ou florestas naturais exceto os sistemas agroflorestais, terras degradadas e inaproveitáveis para a agropecuária), e, portanto, não pertinentes à análise.

O capital foi construído através do logaritmo da soma do valor dos bens, em mil reais, de cada estabelecimento no ano, os quais são: 1) bens imóveis (prédios, instalações e outras benfeitorias, terras, lavouras permanentes e matas plantadas) e 2) outros bens (veículos, tratores, máquinas e implementos, animais para reprodução, criação e outros fins) divididos pela área em hectares.

O IBGE divide em dois tipos o trabalho: O familiar, caracterizado pelo parentesco com o proprietário; e o contratado, ou seja, aqueles que não são familiares. Eles são apresentados no censo pelo número de trabalhadores por classes de dia trabalhados.⁷ Porém, usa-se o logaritmo das horas de trabalho por hectare de modo a incluir os empregados temporários na análise econométrica. Para isto, considera-se 60, 120 e 180 dias trabalhados por ano e que a jornada de trabalho seja de 8 horas diárias. Multiplica-se os dias trabalhados no ano por 8 (horas) e pelo respectivo número de trabalhadores. Os dois tipos de trabalho são somados e tem-se o trabalho total em horas trabalhadas por ano.

⁶ As faixas de áreas são: 1) Mais de 0 a menos de 0,1 ha; 2) De 0,1 a menos de 0,2 ha; 3) De 0,2 a menos de 0,5 há; 4) De 0,5 a menos de 1 ha; 5) De 1 a menos de 2 ha; 6) De 2 a menos de 3 ha; 7) De 3 a menos de 4 ha; 8) De 4 a menos de 5 ha; 9) De 5 a menos de 10 ha; 10) De 10 a menos de 20 ha; 11) De 20 a menos de 50 ha; 12) De 50 a menos de 100 ha; 13) De 100 a menos de 200 ha; 14) De 200 a menos de 500 ha; 15) De 500 a menos de 1.000 ha; 16) De 1.000 a menos de 2.500 ha; 17) De 2.500 ha e mais.

⁷ Menos de 60 dias; de 60 a 180 dias; 180 e mais dias.

Além das observações excluídas descritas acima, excluíram-se as 1% menores e maiores propriedades por serem consideradas *outliers*.⁸ Ademais, para evitar o descarte de observações cujos valores são 0, somou-se 0,001 às variáveis capital e trabalho (Barrett et al., 2010).⁹

Portanto, restaram 5.193 municípios que, considerando todas as informações das 17 faixas de área de cada um, totalizaram 58.862 observações. A Tabela 1 apresenta as estatísticas descritivas da amostra para o Brasil em 2006. Observa-se uma área média de 2.462 hectares com produção média de R\$ 2.450 por hectare (VP/ha). Eles ainda despenderam R\$ 13.930 com capital por hectare e 167,1 horas de mão de obra por hectare. Vale ressaltar que os valores monetários são nominais do ano de 2006. As estatísticas descritivas da amostra Brasil dividida em subamostras para as cinco regiões, também a serem analisadas pelo presente trabalho, estão no Apêndice 1.

Tabela 1 – Estatísticas descritivas da amostra, Brasil.

Variável	Média	Desvio Padrão	Mínimo	Máximo
Valor da Produção/ha (mil reais)	2,45	5,20	0,02	65,56
Área Prod. (ha)	2.462	4.795	2	43.127
Capital/ha (mil reais)	13,93	20,08	0,60	183,30
Trabalho/ha (horas)	167,10	307,3	0,72	2.595

Fonte: Elaborada a partir de dados do Censo Agropecuário de 2006.

5. Resultados

Para investigar a hipótese da relação inversa entre a produtividade e o tamanho das propriedades agrícolas no Brasil partiu-se, primeiramente, de uma função de produção Cobb-Douglas. Conforme discutido na seção anterior, a variável dependente usada nas estimativas paramétricas e não paramétricas é o valor total da produção agrícola (VP) dividido pelo total da terra cultivada em hectares. A escolha da função Cobb-Douglas se dá por ela ser mais parcimoniosa do que uma Translog, evitando assim problemas como a maldição da dimensionalidade presente na econometria *kernel* não paramétrica. Uma outra justificativa para essa escolha é que, embora a função Cobb-Douglas seja mais restritiva do que uma Translog, por exemplo, a primeira tem a vantagem de ser globalmente bem-comportada a partir de uma perspectiva de teoria econômica do que uma Translog (Verschelde et al., 2013; O'Donnell, 2016).

Dada as diferentes regiões do país, uma análise de dados em painel seria mais adequada para controlar a heterogeneidade não observada (ex., condições edafoclimáticas, fatores institucionais) existente entre as propriedades agrícolas e os diferentes municípios. No entanto, devido a problemas técnicos com a recuperação de dados dos Censos Agropecuários de anos anteriores ao Censo 2006, foco do estudo, e pelo Censo Agropecuário 2017 não estar ainda disponível durante a realização desta pesquisa, uma análise mais ampla que contemple uma estrutura de dados longitudinais junto com estimativas de regressões paramétricas e não paramétricas é deixada para estudos futuros. Dessa forma, para remediar esta limitação, optou-se então por comparar as regiões separadamente para verificar se existem grandes diferenças entre elas.

A Tabela 2 mostra os resultados das regressões paramétricas da função de produção Cobb-Douglas estimada por mínimos quadrados ordinários, agregando as 17 faixas de área de cada município pesquisado pelo Censo 2006. Coeficientes que se mostraram estatisticamente significativos ao nível de 10%, 5% ou 1% estão uniformemente destacados em negrito. Enquanto na coluna (1) é apresentado o resultado agregado das 17 faixas de área

⁸ A abordagem não paramétrica pode ser sensível à presença de *outliers*, pois métodos não paramétricos dependem da ponderação local que depende da distância relativa entre as observações (Delgado et al., 2014).

⁹ A robustez das estimativas usando outros números (por exemplo, 0,01) está disponível mediante solicitação com os autores.

de todos os municípios do Brasil, totalizando 58.862 observações, nas demais colunas (2-6) as estimativas dos coeficientes são apresentadas por região começando pela região Sul (coluna 2) até a região Centro-Oeste (coluna 6).

No Brasil (coluna 1), a principal variável de interesse na análise, a área produtiva, apresenta um coeficiente estatisticamente significativo de 0,10. Isto significa que um aumento em 10% da terra cultivada vai gerar um aumento de 1% do valor da produção por área. Conforme atestado por Heltberg (1998), como a medida de produtividade é o valor da produção por hectare, somente valores negativos dos coeficientes área indicariam uma rejeição da hipótese da relação inversa entre o tamanho das propriedades agrícolas e a produtividade.

As maiores elasticidades da terra são observadas na região Sudeste (0,14) e Sul (0,13), e as menores no Norte (0,05), seguida da região Nordeste (0,04). A região Centro-Oeste apresenta uma elasticidade da terra de 0,07. Desta forma, estes coeficientes não negativos e estatisticamente significativos sugerem também que, ao nível regional, não se observou a relação inversa entre produtividade e área em 2006.

Para os demais fatores de produção, capital e trabalho por hectare, as magnitudes dos coeficientes evidenciam uma considerável importância destes sobre o fator terra tendo como base o ano de 2006. Em segundo lugar, observa-se que no Brasil e na maior parte das regiões o capital sobrepõe significativamente sobre a mão de obra. Enquanto o capital apresenta as maiores elasticidades (acima de 0,80) para as regiões Sudeste e Centro-Oeste, o trabalho, para estas regiões supracitadas, apresenta as menores (abaixo de 0,31), por sua vez, uma influência expressiva do primeiro fator em relação ao segundo sobre os ganhos de produtividade, mantendo o fator terra constante. O Norte é a única região cuja elasticidade do trabalho (0,58) é maior que a do capital (0,30). Na região Nordeste, a participação do capital é ligeiramente superior à do trabalho, mas orbitando ambos em torno de 0,45, e mostrando-se igualmente mais importantes para a produtividade na região do que o fator terra.

Tabela 2 – Regressões paramétricas utilizando a função de produção Cobb-Douglas para o Brasil e regiões. Variável dependente: Ln do Valor da Produção/ha

	Brasil	Sul	Sudeste	Norte	Nordeste	Centro-Oeste
	(1)	(2)	(3)	(4)	(5)	(6)
Ln da Área (ha)	0,10	0,13	0,14	0,05	0,04	0,07
	35,04	23,41	25,95	4,81	7,68	6,86
Ln do Capital/ha	0,64	0,67	0,82	0,30	0,48	0,80
	164,75	51,52	82,45	16,03	56,41	40,00
Ln do Trabalho/ha	0,38	0,30	0,31	0,58	0,42	0,24
	105,21	39,15	41,02	42,43	67,55	16,38
Constante	-3,47	-3,22	-3,94	-3,36	-3,07	-3,39
	107,55	48,16	62,19	29,93	56,69	29,62
Observações	58.862	12.557	16.862	4.365	20.511	4.567
R ²	0,55	0,48	0,50	0,63	0,53	0,54

Fonte: Elaborada a partir de dados do Censo Agropecuário de 2006. Nota: 1) Coeficientes estatisticamente significativos ao nível de 10%, 5% ou 1% estão destacados em negrito. 2) Erros padrões robustos para heterocedasticidade estão abaixo dos coeficientes.

Como a estimativa não paramétrica fornece efeitos parciais para cada observação, a Tabela 3 apresenta de forma resumida estes efeitos de cada covariável na forma de três percentis, a saber, 25º(Q1), 50º(Q2) e 75º(Q3), dos três estimadores não paramétricos propostos (LCLS, LLLS, LQLS).¹⁰ O mesmo recorte amostral do Brasil e regiões com valores

¹⁰ Este exercício é análogo à realização de testes *t* de significância de parâmetros em um modelo linear padrão (Delgado et al., 2014).

convertidos em logaritmos como realizado na análise paramétrica da função de produção Cobb-Douglas foi adotado aqui. Nos três estimadores, também, erros padrões dos percentis 25°, 50° e 75° de cada covariável foram calculados através do método conhecido por *wild bootstrap* com 500 replicações. O método *bootstrap* tem como principal vantagem a obtenção de erros padrões que são robustos sobre a presença de heterocedasticidade (Li & Racine, 2007; Henderson & Parmeter, 2015). As larguras das janelas da função *kernel* adotadas, epanechnikov, para todos os estimadores, foram calculadas através do método LSCV descrito na metodologia, seguindo os passos propostos por Henderson & Parmeter (2015). Similar aos resultados paramétricos, efeitos parciais estatisticamente significativos ao nível de 10%, 5% ou 1% estão destacados em negrito.

A primeira observação a ser feita é que o poder explicativo (R^2) dos modelos não paramétricos nos três estimadores adotados para as amostras do Brasil e regiões aumentou em relação aos modelos paramétricos análogos. Em segundo, quando as estimativas dos três fatores de produção de cada percentil (25°, 50° e 75°) são comparadas entre os três estimadores não paramétricos dentro de cada recorte geográfico, os resultados não divergem de forma significativa entre eles, sugerindo que apenas 1 estimador seria suficiente.¹¹ Por fim, enquanto na análise paramétrica, terra, capital e trabalho apresentaram sinal positivo ao nível de Brasil e regiões, na análise não paramétrica as estimativas destes três insumos são muito mais variáveis entre os percentis, evidenciando que possíveis erros de especificação na forma funcional a ser estimada, que não levem em conta a verdadeira estrutura subjacente dos dados, podem conduzir a resultados equivocados.

No caso do Brasil (linhas 1-3) com um todo, a variável de interesse, área, apresenta percentis positivos e estatisticamente significativos na mediana (50°) para o estimador LCLS (elasticidade de 0,27), e no terceiro quartil (75°) para os estimadores LCLS (elasticidade de 1,01) e LLLS (elasticidade de 1,23). Portanto, semelhante aos resultados encontrados na análise paramétrica, e considerando somente os percentis que foram estatisticamente significativos, a relação inversa entre área e produtividade também não é observada quando testada não parametricamente no âmbito nacional.

A magnitude dos efeitos parciais de capital e trabalho por hectare apresentam uma relação direta com o aumento do percentil sobre o impacto na produtividade, conforme o esperado. No entanto, observa-se uma grande distância do efeito destes dois insumos entre os produtores dos menores percentis (25°) e dos maiores (75°). Por exemplo, no estimador LLLS, a elasticidade do fator trabalho oscila entre -1,02 (25°) e 2,99 (75°), enquanto do capital entre -1,27 (25°) e 2,70 (75°). Nota-se, também, para o terceiro quartil 75° deste estimador, que o capital apresenta elasticidade muito maior (2,70) do que com relação à área (1,23), ou seja, um aumento no capital destas propriedades gera mais que o dobro de aumento de produtividade do que o uso da área. O fator trabalho apresenta efeitos parciais tanto em magnitude de valores quanto em tendência dos sinais, muito semelhantes ao fator capital nos três estimadores adotados.

Na análise regional (linhas 4-18 da Tabela 3), com exceção da região Sul, a hipótese da relação inversa é corroborada no percentil 25° (propriedades na faixa de área de até 100, 200 ou 500 hectares dependendo da região) em quase todos os estimadores. Este fato pode sugerir que esforços de políticas (ex., apoio público à pesquisa e extensão agropecuária) destinados a aumentar a produtividade agrícola e o crescimento da produção entre pequenos agricultores poderiam ser mais eficazes no crescimento da produção do que a adoção de uma política de redistribuição de terras (Lipton, 2009). A hipótese é uma avenida promissora para estudos futuros. Observa-se, também, que as elasticidades da terra que se mostram negativas e estatisticamente significativas em algum dos três estimadores no percentil 25° na região Centro-Oeste (produtores de até 500 ha) chegam a ser bem mais inferiores (abaixo de -1) do que as elasticidades no mesmo percentil da região Sudeste e Nordeste (produtores de até 100 ha) e Norte (produtores de até 200 ha).

¹¹ Segundo Henderson & Parmeter (2015), o uso do estimador LLLS é mais recomendado por oferecer recursos teóricos e aplicados mais desejáveis que os estimadores LCLS e LQLS.

Nos percentis superiores, na mediana, por exemplo, a relação negativa é observada entre área e valor da produção por hectare nas propriedades que possuem entre 500 e 2500 hectares apenas na região Centro-Oeste, mas não nas outras regiões. No percentil superior de 75°, a hipótese da relação inversa não é corroborada em nenhuma das regiões do país. Ademais, neste percentil, efeitos parciais da terra positivos e estatisticamente significativos são observados nas regiões Sul, Nordeste e Sudeste (propriedades acima de 500 ha) e Norte (propriedades acima de 1000 ha).

Os resultados não paramétricos aqui apresentados parecem não divergir completamente dos encontrados na literatura sobre o Brasil, com algumas diferenças principalmente se a definição das pequenas e médias propriedades é levada em consideração. Helfand & Levine (2004), utilizando dados do Censo Agropecuário de 1995/1996 e análise envoltória de dados (DEA) para a região Centro-Oeste, mostraram que a relação entre tamanho e eficiência da fazenda foi considerada não linear, com a eficiência diminuindo primeiro e depois aumentando com o tamanho. Para fazendas de cerca de 1000 a 2000 ha, a eficiência caiu à medida que o tamanho da fazenda aumentou, mas além desse tamanho, começou a aumentar novamente. Os autores sugerem que o tipo de posse da terra, o acesso a instituições e insumos modernos são importantes determinantes das diferenças de eficiência entre as fazendas.

Mais recentemente, Rada et al. (2019), utilizando fronteira estocástica e microdados dos Censos Agropecuários para os anos de 1985 a 2006, divididos em 5 classes de tamanhos das propriedades, observaram uma relação em formato de U, a qual as menores (0-5 hectares) e as maiores (500 hectares e mais) propriedades tiveram maior crescimento da PTF do que as medianas. Segundo os autores, os resultados variaram bastante de acordo com as regiões e o grau de especialização do produtor, o que evidencia sobremaneira a heterogeneidade da agropecuária brasileira e a importância da análise regionalizada. Eles também encontraram que a educação está associada a um crescimento da produtividade mais rápido independentemente do tamanho da fazenda, enquanto efeitos positivos da assistência técnica e negativos do crédito rural sobre a produtividade estavam associadas às propriedades de tamanho maiores, sendo este último um resultado robusto, mas não esperado segundo os autores. Já Lázari & Magalhães (2016), também utilizando dados do Censo Agropecuários entre 1985 e 2006 da região Sudeste, com o recorte de 0-5 ha, 5-20 ha, 20-100 ha, 100-500 ha e 500 ha ou mais, também não corroboraram a relação inversa ao analisar o crescimento da PTF por área.

As estimativas não paramétricas de capital e trabalho das regiões apresentaram resultados de elasticidades muito similares em magnitudes e sinais quando comparados com os resultados do Brasil nos três estimadores. A região Centro-Oeste é uma exceção. Nota-se, também, que para cada estimador, variações observadas entre os percentis 25° e 75° do capital por área são significativamente mais discrepantes do que as variações do fator trabalho no mesmo intervalo de percentis nas regiões Sudeste e Sul. Se o fator capital tem um peso maior do que o fator trabalho nestas duas regiões, por outro lado, a resposta do fator trabalho por área sobre a produtividade agrícola é mais importante nas regiões Norte e Nordeste do que o capital, conforme observado na análise paramétrica.

Como mencionado, ao contrário das quatro regiões acima, os resultados não paramétricos da região Centro-Oeste não corroboram completamente o cenário nacional, mostrando inclusive alguns resultados bastante contraintuitivos se comparados com a análise paramétrica e com a literatura. Enquanto a análise paramétrica apresenta elasticidades positivas para capital e trabalho, conforme o esperado, elasticidades de capital e trabalho negativas ou mesmo próximas de zero foram observadas na análise não paramétrica, principalmente no percentil 75°. Estes resultados discrepantes podem estar sugerindo que outros fatores, como educação, acesso a mercados, aspectos institucionais e ações envolvendo políticas macroeconômicas e políticas setoriais específicas, podem ter assumido um papel mais importante para ganhos de produtividade em 2006 do que somente o uso intensivo da terra, capital e trabalho na região (Helfand & Levine, 2004, Gasques et al., 2014, Rada et al., 2019).

Tabela 3 – Regressões não paramétricas para diferentes estimadores selecionados para o Brasil e regiões. Variável dependente: Ln do Valor da Produção/ha

	Ln da Área (ha)			Ln do Capital/ha			Ln do Trabalho/ha			R ²
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	
Brasil										
LCLS (1)	-0,24	0,27	1,01	-1,19	0,03	2,58	-0,98	0,005	2,58	0,59
	-0,67	1,66	0,20	0,14	0,01	0,49	0,25	-0,01	0,09	
LLLS (2)	-0,40	0,28	1,23	-1,27	0,06	2,70	-1,02	0,09	2,99	0,59
	-0,33	-0,19	0,30	0,20	-0,05	0,29	0,24	-0,39	0,85	
LQLS (3)	-0,55	0,25	1,20	-1,28	0,02	2,53	-1,06	0,01	2,86	0,60
	-0,61	-0,36	-0,85	-0,87	0,06	0,66	0,41	0,006	0,20	
Sul										
LCLS (4)	0,03	0,38	0,73	0,49	2,15	3,23	-0,11	0,97	1,76	0,52
	-0,49	-1,28	-2,66	-0,81	0,90	0,34	0,07	0,50	0,58	
LLLS (5)	0,04	0,62	1,20	0,62	2,35	3,91	-0,08	1,25	2,26	0,53
	-0,09	0,33	0,28	-1,15	0,18	0,24	0,04	0,29	0,51	
LQLS (6)	-0,09	0,55	1,23	0,22	2,17	3,70	-0,22	1,28	2,37	0,54
	-0,16	0,11	0,08	-0,70	0,27	0,87	0,07	0,19	1,08	
Sudeste										
LCLS (7)	-0,54	0,34	1,40	-2,34	1,30	3,88	-0,66	0,81	2,51	0,55
	0,31	-1,22	0,24	0,32	0,45	0,73	0,06	0,47	0,30	
LLLS (8)	-0,81	0,41	1,72	-2,48	1,43	3,87	-0,65	0,94	2,89	0,55
	0,46	0,16	0,171	0,09	0,26	0,38	0,04	0,06	0,45	
LQLS (9)	-0,93	0,39	1,64	-2,52	1,30	3,93	-0,71	0,65	2,65	0,56
	0,50	-0,28	0,13	0,41	0,24	0,43	0,08	-0,88	0,38	
Norte										
LCLS (10)	-0,34	-0,03	0,34	-0,33	-0,10	0,20	-1,43	-0,11	3,24	0,68
	-0,26	-0,11	0,14	0,09	0,01	-0,19	0,14	0,034	1,80	
LLLS (11)	-0,47	-0,03	0,41	-0,61	-0,17	0,27	-1,66	-0,14	3,75	0,68
	0,24	-0,30	-0,38	0,10	0,03	-0,43	0,13	0,02	0,63	
LQLS (12)	-0,60	-0,04	0,52	-0,59	-0,14	0,27	-1,54	-0,16	3,29	0,70
	0,31	-0,12	0,10	0,04	0,05	-3,87	0,20	0,02	1,27	
Nordeste										
LCLS (13)	-0,06	0,23	0,79	-0,73	-0,10	0,93	-1,83	-0,47	2,76	0,57
	-0,14	-0,30	0,11	-0,57	0,003	-0,76	0,48	0,04	0,80	
LLLS (14)	-0,23	0,24	0,99	-0,73	-0,09	0,98	-1,76	-0,45	3,55	0,57
	-0,20	-0,16	-0,81	0,04	0,02	-0,79	0,23	0,07	0,64	
LQLS (15)	-0,70	0,06	0,86	-2,06	-0,33	2,75	-1,81	-0,45	1,50	0,62
	0,29	-0,10	-1,41	0,30	0,02	0,52	0,12	0,07	0,20	
Centro-Oeste										
LCLS (16)	-0,61	-0,18	0,15	-2,27	-1,31	-0,22	-0,93	-0,08	0,03	0,60
	-0,89	-0,17	-1,74	0,15	0,30	-1,11	0,18	-0,05	0,02	
LLLS (17)	-1,27	-0,51	-0,01	-2,66	-1,37	-0,32	-1,23	-0,13	0,008	0,60
	0,35	0,14	-0,15	0,16	0,08	0,05	0,14	0,08	0,003	
LQLS (18)	-1,63	-0,56	0,10	-2,51	-1,28	-0,04	-1,37	-0,20	0,01	0,62
	0,70	0,34	-0,16	0,50	0,10	-0,07	-1,12	0,05	-0,04	

Fonte: Elaborada a partir de dados do Censo Agropecuário de 2006. Nota: 1) Tabela mostra efeitos parciais ao nível percentil de 25° (Q1), 50° (Q2), e 75° (Q3). 2) Efeitos parciais estatisticamente significativos ao nível de 10%, 5% ou 1% estão destacados em negrito. 3) Erros padrões robustos para heterocedasticidade estão abaixo dos efeitos parciais.

6. Considerações Finais

O uso de funções de produção para analisar a relação entre insumos e produtos tem sido feito há anos. No entanto, a decisão sobre a sua forma funcional a ser estimada é de suma importância porque implica suposições a priori à estimação dos dados no modelo (Ulveling & Fletcher, 1970).

Motivado por Henderson & Parmeter (2015), que propõem o uso da análise de regressão *kernel* não paramétrica para evitar vieses das formas funcionais e aconselham o uso dela na estimação de funções de produção, este trabalho procurou testar se existe uma relação inversa entre produtividade e área para o Brasil através dos dados do Censo Agropecuário de 2006 executado pelo IBGE. A estratégia empregada consistiu em comparar a análise paramétrica, cujo estimador “enxerga” os dados de uma forma global com a não paramétrica *kernel*, que analisa os dados localmente, mas que ainda não foi utilizada para este problema de pesquisa no Brasil. Além disso, até hoje, aplicações de métodos de regressão não paramétricos, embora largamente discutidas na literatura internacional, ainda não foram bem documentadas no país e, portanto, merecem mais atenção em trabalhos futuros.

Enquanto a análise paramétrica mostrou que a relação é direta entre produtividade e área cultivada no âmbito nacional e regional, a não paramétrica explicitou, por sua vez, que a hipótese da relação inversa entre a área e a produtividade foi confirmada para as menores propriedades, porém rejeitada para as médias e grandes, mas apenas quando regiões são analisadas separadamente.

A concentração de terras tem mudando pouco no Brasil e continua ainda muito alta (Hoffmann & Ney, 2010; Ferreira-Filho & Vian, 2016). Em 2006, a parcela de terra ocupada pelas 5% maiores fazendas era quase 30 vezes maior que a parcela ocupada pelas 50% menores propriedades com um índice de Gini na casa de 0.85 (Hoffmann & Ney, 2010). No entanto, com relação às propostas de políticas relacionadas, do ponto de vista da produtividade, uma redistribuição de terras talvez não seria tão justificada entre as propriedades das diferentes regiões nos percentis de área cuja relação foi inversa, porque, como já sugerido pelo presente trabalho, uma distribuição de terras também deve ser acompanhada de intervenções políticas destinadas a melhorar o capital social e humano, juntamente com outros mecanismos que promovam melhorias na produtividade e uso da terra (World Bank, 2008; Chen et al., 2011). Ademais, redução das distorções, principalmente nos mercados de trabalho, terra e crédito, são iniciativas difíceis de serem implementadas, contudo, contribuem positivamente para melhoraria da renda e condições de vida entre os pequenos produtores em países em desenvolvimento (Julien et al., 2019).

Para trabalhos futuros, o uso de um nível de desagregação por fazenda (ao nível de microdados) poderia trazer ganhos adicionais de robustez principalmente na análise não paramétrica. Além disso, a adoção de dados em painel atualmente tem feito parte do “estado da arte” nas análises econométricas, permitindo o controle de características não observadas e invariantes no tempo das regiões e produtores, trazendo também vantagens substanciais na análise. Assim, a 1) falta de dados longitudinais mais à 2) maldição de dimensionalidade embutidas nas regressões não paramétricas que pode ser corrigida pela adoção de modelos semiparamétricos que permitem a inclusão de outras covariáveis igualmente importantes na função da produção como o nível de escolaridade do produtor, idade e sexo, se recebeu treinamento técnico, etc., constituíram-se ambas como limitações do presente estudo.

Agradecimentos

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Referências

- Ali, A. D., & Deininger, K. (2015). Is there a farm size-productivity relationship in African agriculture? Evidence from Rwanda. *Land Economics*, 91(2), 317-343. <http://dx.doi.org/10.3368/le.91.2.317>
- Assunção, J. J., & Braido, L. H. B. (2007). Testing household-specific explanations for the inverse productivity relationship. *American Journal of Agricultural Economics*, 89(4), 980-990. <http://dx.doi.org/10.1111/j.1467-8276.2007.01032.x>
- Assunção, J. J., & Ghatak, M. (2003). Can unobserved heterogeneity in farmer ability explain the inverse relationship between farm size and productivity. *Economics Letters*, 80(3), 189-194. [http://dx.doi.org/10.1016/S0165-1765\(03\)00091-0](http://dx.doi.org/10.1016/S0165-1765(03)00091-0)

- Ávila, A. F. D., Garagorry, F., & Cardoso, C. C. (2013). Produção e produtividade da agricultura brasileira: taxas de crescimento, comparações regionais e seus determinantes. In E. R. A. Alves, G. S. Souza & E. G. Gomes. *Contribuição da Embrapa para o desenvolvimento da agricultura no Brasil* (pp. 87-124). Brasília: Embrapa.
- Bardhan, P. K. (1973). Size, productivity, and returns to scale, An analysis of farm-level data in Indian agriculture. *Journal of Political Economy*, 81(6), 1370-1386. <http://dx.doi.org/10.1086/260132>
- Barrett, C. B. (1996). On price risk and the inverse farm size-productivity relationship. *Journal of Development Economics*, 51(2), 193-215. [http://dx.doi.org/10.1016/S0304-3878\(96\)00412-9](http://dx.doi.org/10.1016/S0304-3878(96)00412-9)
- Barrett, C. B., Bellemare, M. F., & Hou, J. Y. (2010). Reconsidering conventional explanations of the inverse productivity-size relationship. *World Development*, 38(1), 88-97. <http://dx.doi.org/10.1016/j.worlddev.2009.06.002>
- Benjamin, D. (1995). Can unobserved land quality explain the inverse productivity relationship? *Journal of Development Economics, Amsterdam*, 46(1), 51-84. [http://dx.doi.org/10.1016/0304-3878\(94\)00048-H](http://dx.doi.org/10.1016/0304-3878(94)00048-H)
- Berry, R. A., & Cline, W. R. (1979). *Agrarian structure and productivity and developing countries*. Baltimore: Johns Hopkins University Press.
- Bhalla, S. (1988). Does quality matter? Theory and measurement. *Journal of Development Economics*, 29(1), 45-62. [http://dx.doi.org/10.1016/0304-3878\(88\)90070-3](http://dx.doi.org/10.1016/0304-3878(88)90070-3)
- Binswanger, H. P., & Rosenzweig, M. R. (1986). Behavioural and material determinants of production relations in agriculture. *The Journal of Development Studies*, 22(3), 503-539. <http://dx.doi.org/10.1080/00220388608421994>
- Binswanger, H. P., Deininger, K., & Feder, G. (1995). Power, distortions, revolt and reforming agricultural land relations. In J. Behrman & T. N. Srinivasan (Eds.), *Handbook of development economics* (Vol. 3B, Cap. 42, p. 2659-2772). Amsterdam: North-Holland Publishing.
- Blundell, R., & Duncan, A. (1998). Kernel regression in empirical economics. *The Journal of Human Resources*, 3, 435-461.
- Bragagnolo, C., & Barros, G. S. C. (2015). Impactos dinâmicos dos fatores de produção e da produtividade sobre a função de produção agrícola. *Revista de Economia e Sociologia Rural*, 53(1), 31-50. <http://dx.doi.org/10.1590/1234-56781806-9479005301002>
- Bragagnolo, C., Spolador, H. F. S., & Barros, G. S. C. (2010). Regional Brazilian agriculture TFP analysis, a stochastic frontier analysis approach. *Economía*, 11(4), 217-242.
- Brasil. (2016, maio 12). Medida provisória nº 726, de 12 de maio de 2016. *Diário Oficial [da] República Federativa do Brasil*, Brasília, seção 1, p. 1.
- Carletto, C., Savastano, S., & Zezza, A. (2013). Fact or artifact, the impact of measurement errors on the farm size-productivity relationship. *Journal of Development Economics*, 103, 254-261. <http://dx.doi.org/10.1016/j.jdeveco.2013.03.004>
- Carrer, M. J., & Souza Filho, H. M. (2016). Economias de escala e eficiência econômica na produção de laranja no estado de São Paulo. *Revista de Economia e Sociologia Rural*, 54(1), 51-70. <http://dx.doi.org/10.1590/1234-56781806-9479005401003>
- Carter, M. R. (1984). Identification of the inverse relationship between farm size and productivity, an empirical analysis of peasant agricultural productions'. *Oxford Economic Papers*, 36(1), 131-145. <http://dx.doi.org/10.1093/oxfordjournals.oep.a041621>
- Carter, M. R., & Wiebe, K. (1990). Access to capital and its impact on agrarian structure and productivity in Kenya. *American Journal of Agricultural Economics*, 72(5), 1146-1150. <http://dx.doi.org/10.2307/1242523>
- Chayanov, A. V. (1926). *The theory of peasant economy* (D. Thorner, B. Kerblay & R. E. F. Smith, eds.). Homewood: Irwin Publishers. Recuperado em 11 de outubro de 2017, de https://growthecon.com/assets/papers/alexander_chayanov_the_theory_of_peasant_economy.pdf
- Chen, Z., Huffman, W. E., & Rozelle, S. (2011). Inverse relationship between productivity and farm size, the case of China. *Contemporary Economic Policy, Morgantown*, 29(4), 580-592. <http://dx.doi.org/10.1111/j.1465-7287.2010.00236.x>
- Deaton, A. (1989). Rice prices and income distribution in Thailand, a non-parametric analysis. *Economic Journal*, 99(395), 1-37. <http://dx.doi.org/10.2307/2234068>
- Deininger, K., & Byerlee, D. (2012). The rise of large farms in land abundant countries, do they have a future? *World Development*, 40(4), 701-714. <http://dx.doi.org/10.1016/j.worlddev.2011.04.030>

- Delgado, M. S., Henderson, D. J. & Parmeter, C. F. (2014). Does education matter for economic growth? *Oxford Bulletin of Economics and Statistics*, 76(3), 334-359. <http://dx.doi.org/10.1111/obes.12025>.
- Eastwood, R., Lipton, M., & Newell, A. (2010). Farm size. In R. Evenson & P. Pingali (Eds.), *Handbook of agricultural economics* (Vol. 4, Cap. 65, pp. 3323-3397). Elsevier.
- Eubank, R. L. (1999). *Nonparametric regression and spline smoothing* (2nd ed.). Nova Iorque: Marcel Dekker. <http://dx.doi.org/10.1201/9781482273144>.
- Feder, G. (1985). The relation between farm size and farm productivity, the role of family labor, supervision and credit constraints. *Journal of Development Economics*, 18(2), 297-313. [http://dx.doi.org/10.1016/0304-3878\(85\)90059-8](http://dx.doi.org/10.1016/0304-3878(85)90059-8)
- Felama, J., Raiher, A. P., & Ferreira, C. R. (2013). Agropecuária brasileira, desempenho regional e determinantes de produtividade. *Revista de Economia e Sociologia Rural*, 51(3), 555-573. <http://dx.doi.org/10.1590/S0103-20032013000300008>
- Ferreira, C. B., Araujo, J. A., Tabosa, F. S., & Lima, J. R. F. (2016). Produtividade Agrícola nos Países da América Latina. *Revista de Economia e Sociologia Rural*, 54(3), 437-458. <http://dx.doi.org/10.1590/1234-56781806-94790540303>
- Ferreira-Filho, J. B., & Vian, C. E. F. (2016). The evolving role of large and medium farms on Brazilian Agriculture. *Agricultural Economics*, 47(S1), 215-225. <http://dx.doi.org/10.1111/agec.12310>
- Gasques, J. G., Bastos, E. T., Valdes, C., & Bacchi, M. R. P. (2012). Produtividade da agricultura brasileira e os efeitos de algumas políticas. *Revista de Política Agrícola*, 21(3), 83-92.
- Gasques, J. G., Bastos, E. T., Valdes, C., & Bacchi, M. R. P. (2014). Produtividade da agricultura, resultados para o Brasil e estados selecionados. *Revista de Política Agrícola*, 23(3), 87-98.
- Helfand, S. M., & Levine, E. S. (2004). Farm size and the determinants of productive efficiency in the Brazilian Center-West. *Agricultural Economics*, 31(2), 241-249. <http://dx.doi.org/10.1111/j.1574-0862.2004.tb00261.x>
- Helfand, S. M., Magalhães, M. M., & Rada, N. E. (2015). *Brazil's agricultural total factor productivity growth by farm size* (Inter-American Development Bank Working Paper Series, No. 609). Washington: IADB. Recuperado em 07 de julho de 2020, de <https://publications.iadb.org/publications/english/document/Brazil-Agricultural-Total-Factor-Productivity-Growth-by-Farm-Size.pdf>
- Heltberg, R. (1998). Rural market imperfections and the farm size-productivity relationship, evidence from Pakistan. *World Development, Amsterdam*, 26(10), 1807-1826. [http://dx.doi.org/10.1016/S0305-750X\(98\)00084-9](http://dx.doi.org/10.1016/S0305-750X(98)00084-9)
- Henderson, D. J., & Parmeter, C. F. (2015). *Applied nonparametric econometrics*. Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511845765>.
- Henderson, H. (2015). Considering technical and allocative efficiency in the inverse farm size-productivity relationship. *Journal of Agricultural Economics*, 66(2), 442-469. <http://dx.doi.org/10.1111/1477-9552.12086>
- Hoffmann, R., & Ney, M. G. (2010). *Estrutura fundiária e propriedade agrícola no Brasil, grandes regiões e unidades da federação*. Brasília: Ministério do Desenvolvimento Agrário. Recuperado em 11 de outubro de 2017, de http://nead.mda.gov.br/download.php?file=publicacoes/outras/estrutura_fundiaria_e_propriedade_agricola_no_brasil.pdf
- Instituto Brasileiro de Geografia e Estatística – IBGE. (2016). *Sistema IBGE de Recuperação Automática – SIDRA*. Rio de Janeiro: IBGE. Recuperado em 11 de outubro de 2017, de [https://sidra.ibge.gov.br/pesquisa/censo-agropecuaria/censo-agropecuario-2006/segunda-apuracao\(Eds.\)](https://sidra.ibge.gov.br/pesquisa/censo-agropecuaria/censo-agropecuario-2006/segunda-apuracao(Eds.))
- Julien, J. C., Bravo-Ureta, B. E., & Rada, N. E. (2019). Assessing farm performance by size in Malawi, Tanzania, and Uganda. *Food Policy*, 84, 153-164. <http://dx.doi.org/10.1016/j.foodpol.2018.03.016>
- Lamb, R. L. (2003). Inverse productivity, land quality, labor markets, and measurement error. *Journal of Development Economics*, 71(1), 71-95. [http://dx.doi.org/10.1016/S0304-3878\(02\)00134-7](http://dx.doi.org/10.1016/S0304-3878(02)00134-7)
- Lázari, N. C., & Magalhães, M. M. (2016). Crescimento da PTF segundo tamanho de estabelecimentos rurais na região Sudeste, de 1985 a 2006. *Revista de Economia e Sociologia Rural*, 57(2), 198-214. <http://dx.doi.org/10.1590/1806-9479.2019.175194>
- Li, Q., & Racine, J. S. (2007). *Nonparametrics econometrics, theory and practice*. Princeton: Princeton University Press.

- Lipton, M. (2009). *Land Reform in Developing Countries, Property Rights and Property Wrongs*. New York: Routledge. <http://dx.doi.org/10.4324/9780203876251>.
- Magalhães, M. M., Souza Filho, H. M., Sousa, M. R., Silveira, J. M. F. J., & Buainain, A. M. (2011). Land reform in NE Brazil, a stochastic frontier production efficiency evaluation. *Revista de Economia e Sociologia Rural*, 49(1), 9-30. <http://dx.doi.org/10.1590/S0103-20032011000100001>
- Mendes, M. S., Texeira, C. E., & Salvato, M. A. (2009). Investimentos em infraestrutura e produtividade total dos fatores na agricultura brasileira, 1985-2004. *Revista Brasileira de Economia*, 63(2), 91-102.
- Nadaraya, E. A. (1964). On estimating regression. *Teoriya Veroyatnostei i ee Primeneniya*, 9, 157-159.
- O'Donnell, C. J. (2016). Using information about technologies, markets and firm behavior to decompose a proper productivity index'. *Journal of Econometrics*, 190(2), 328-340. <http://dx.doi.org/10.1016/j.jeconom.2015.06.009>
- Rada, N., & Buccola, S. T. (2012). Agricultural policy and productivity, evidence from Brazilian censuses. *Agricultural Economics*, 43(4), 355-367. <http://dx.doi.org/10.1111/j.1574-0862.2012.00588.x>
- Rada, N., & Fuglie, K. O. (2019). New perspectives on farm size and productivity. *Food Policy*, 84, 147-152. <http://dx.doi.org/10.1016/j.foodpol.2018.03.015>
- Rada, N., & Valdes, C. (2012). *Policy, technology and efficiency of Brazilian agriculture* (No. 137, pp. 1-30). Economic Research Service.
- Rada, N., Helfand, S., & Magalhães, M. (2019). Agricultural productivity growth in Brazil, large and small farmers excel. *Food Policy*, 84, 176-185. <http://dx.doi.org/10.1016/j.foodpol.2018.03.014>
- Sen, A. K. (1962). An aspect of Indian agriculture. *The Economic Weekly*, 14(4-6), 243-246.
- Sen, A. K. (1975). *Employment, technology, and development*. Oxford: Clarendon Press.
- Soares, P., & Spolador, H. F. S. (2019). Eficiência técnica da produção de milho no estado de São Paulo, uma abordagem por metafronteira estocástica. *Revista de Economia e Sociologia Rural*, 57(4), 545-558.
- Uveling, F. E., & Fletcher, L. B. (1970). A Cobb-Douglas production function with variable returns to scale. *American Journal of Agricultural Economics*, 52(2), 322-326. <http://dx.doi.org/10.2307/1237508>
- Verschelde, M., D'Haese, M., Rayp, G., & Vandamme, E. (2013). Challenging small-scale farming, a non-parametric analysis of the (inverse) relationship between farm productivity and farm size in Burundi. *Journal of Agricultural Economics*, 64(2), 319-342. <http://dx.doi.org/10.1111/j.1477-9552.2012.00373.x>
- Vicente, J. R., Anéfalos, L. C., & Caser, D. V. (2000). Produtividade agrícola no Brasil, 1970-1995. In 38º Congresso Brasileiro de Economia e Sociologia Rural. Rio de Janeiro.
- Watson, G. S. (1964). Smooth regression analysis. *Sankhay*, 26, 175-184.
- World Bank. (2008). *World development report 2008, agriculture for development*. Washington, DC: The World Bank Press.
- Yatchew, A. (2003). *Semiparametric regression for the applied econometrician*. Cambridge, Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511615887>.
- Yotopoulos, P. A., & Lau, L. J. (1973). A test for relative economic efficiency, some further results. *The American Economic Review*, 63, 214-223.

Submetido: 17/maio/2019

Aceito: 15/out/2020

Classificação JEL: D24, Q12, Q15, C14

Apêndice 1. Estatísticas descritivas das subamostras, Regiões.

	Média	Desvio Padrão	Mínimo	Máximo
Norte				
Valor da Produção/ha (mil reais)	2,32	5,81	0,023	63,33
Área Prod. (ha)	4.330	7.311	2	42.962
Capital/ha (mil reais)	7,09	13,08	0,60	174
Trabalho/ha (horas)	176,70	350,50	0,72	2.580
Nordeste				
Valor da Produção/ha (mil reais)	1,821	4,29	0,02	64,14
Área Prod. (ha)	1.899	3.551	2	43124
Capital/ha (mil reais)	5,89	9,87	0,60	181,4
Trabalho/ha (horas)	222,20	368,40	0,72	2.595
Centro-Oeste				
Valor da Produção/ha (mil reais)	1,42	3,85	0,02	64,58
Área Prod. (ha)	5.335	8.457	2	43.127
Capital/ha (mil reais)	11,76	17,60	0,66	178,70
Trabalho/ha (horas)	86,55	181,10	0,72	2.440
Sudeste				
Valor da Produção/ha (mil reais)	2,75	5,68	0,02	65,33
Área Prod. (ha)	2.110	3.936	2	41.752
Capital/ha (mil reais)	19,38	23,27	0,62	181,80
Trabalho/ha (horas)	130,70	254,30	0,72	2.590
Sul				
Valor da Produção/ha (mil reais)	3,48	5,85	0,02	65,56
Área Prod. (ha)	2.158	4.100	2	42.883
Capital/ha (mil reais)	22,88	24,27	0,61	183,30
Trabalho/ha (horas)	152,20	264,90	0,72	2.586

Fonte: Elaborada a partir de dados do Censo Agropecuário de 2006.