Monitoring South Africa's metropolitan economies: A survey of the data landscape

Discussion Document 13

MAY 2023

Dieter von Fintel

About ERSA's Discussion Documents

Discussion documents are generally solicited pieces on topical issues of relevance to the national economic debate. The intention is to provide a summary of the issue, accompanied by a discussion about its relevance, importance, and way forward in South Africa. Generally, these are narrative driven contributions, relying on existing work and high-level analysis.

We provide the opportunity for contribution from all relevant perspectives, and therefore these papers do not represent a position by ERSA, its associates, or funders on the identified issues.

We hope that through this we can contribute to a more constructive and informed economic debate. We are particularly interested in hearing your thoughts and comments on these contributions. Please feel free to contact us directly or through LinkedIn. If you feel that you have a contribution that you would like to be part of this series, please contact us directly at research@econrsa.org

Matthew Simmonds
Director

Monitoring South Africa's metropolitan economies: A survey of the data landscape[1,2]

Abstract

Data and local economic development have the potential to be symbiotic. Local administrations in thriving parts of the economy automatically collect extensive data as a product of managing local municipalities and metropoles. However, data creation also offers a tool for better planning and implementation, with the potential for supporting and reinforcing economic development patterns in these wealthier regions. Lagging regions do not have the same data generation capacity, adding a constraint to their ability to plan and monitor local development initiatives. Variations in data infrastrucures across local administrations therefore produces information of disparate quality. Importantly, outputs are not harmonised, complicating comparative planning efforts. Rapid urbanisation means that economic growth in South Africa has become increasingly concentrated in metropolitan areas. Understanding how these local economies develop is essential for interpreting aggregate trends and for contextualising changing patterns of spatial inequality between urban and rural areas. Analysts have long lamented the significant gaps in regional data availability in South Africa, and the situation has been slow to change (Coetzee, 2012; Bezuidenhout et al., 2021). Demand for regional and metropolitan data continues to exceed its supply. Using prior data inventories as a point of departure, this paper is the first to compare trends estimated from a range of survey, administrative and private sector sources. These include the census, model-based population estimates, official StatsSA employment statistics, labour market data from the SARS spatialised panel, widely consulted data harmonised Quentec, and unconventional sources such as night lights luminosity recorded by satellites. The analysis shows that while there are not yet agreeable "gold standards", the options for credible analysis are on the uptick. Up to date local population data rely on updating the latest census using parametric or demographic assumptions. Reliable population statistics would benefit from more regular census enumerations, but leveraging daytime satellite data may be a realistic alternative. Importantly, all household survey data are calibrated to population estimates, so that

[1] Dieter.vonfintel@gmail.com

[2] Acknowledgments: This paper is funded by Economic Research Southern Africa and the Centre for Development Enterprise. I gratefully acknowledge useful comments from Ed Glaeser, Antony Altbeker, Justin Visagie and Diego Santa Maria. Errors remain my own.

producing up-to-date, geographically granular population statistics should enjoy priority. While the sources provide different answers to how many people are employed in various metropolitan areas, the dynamics are remarkably comparable. Statistics on remuneration and local-level GDP are much harder to come by because of a range of biases. From the perspective of high time frequency sources, household surveys suffer from respondent error, post-processing procedures that are undocumented and calibration to outdated sampling frames. In periods of rapid change, sampling frames should be refreshed more regularly. The recent release of spatially-explicit administrative tax data from SARS presents new opportunities for monitoring metropolitan economies: median earnings from this source follow similar trajectories to night lights luminosity data. Apart from some headquarter bias in employment some metropoles, the SARS spatial panel appears to capture formal sector metropolitan employment reliably compared to other sources. Deriving output statistics from this source is therefore promising avenue for further research. Data privately processed and harmonised by Quantec is useful from the perspective of methodological coherence, but the credibility of the source would improve if the post-processing method- ology were transparently documented. Considering three criteria for regional metropolitan data, this paper concludes that availability is on the uptick, transparent methodology should be encouraged, and more research is required to verify data “gold standards”.

Keywords: Metropolitan Statistics; Urbanisation; Population; Employment; Wages; Local Economic Activity; Night Lights

JEL Classifications: R23; R58; H7; C89

1. Introduction

Economic activity and populations are becoming more concentrated in urban areas across the globe. South Africa is no exception. In the 1970s only 20% of South Africans lived in urban areas. By 2018, almost 60% of GDP, half of employment and 40% of people were located in South Africa’s eight metropolitan areas (City Support Programme, 2018). The policy environment has not adapted fast enough to these rapid changes to the economic and demographic landscape (Turok, 2021). In particular, urban service delivery is under pressure, with demand growing faster than supply, and urban labour markets are becoming less efficient. Despite the slow response, urbanfocussed policy is nonetheless gaining recognition for its importance in broader development debates. For these reasons, Integrated Development Planning (IDP) is now mandated to improve municipal budgeting and planning in South Africa. The demand for urban statistics – especially in metropolitan areas – has grown in tandem. A range of cross-sectoral initiatives have been launched to build new data resources and improve capacity for urban planning. For instance, in 2008 the Gauteng City Region established an "Observatory" in partnership with universities that collects sub-municipal data and increases analytical capacity on urban policy (Gauteng City-Region Observatory, 2008); the Western Cape Government has produced annual Municipal Economic Reviews since 2012 (Western Cape Government, 2022); eThekwini municipality monitors ward-level GDP (Matarutse, 2021) and the non-profit South African Cities Network produces annual reports on the state of cities, and has built an urban data repository (Mudau, 2019). These developments acknowledge the importance of data for planning localised budgets and development projects. Furthermore, public discourse is also focused on urban priorities. Spatial inequality – defined here by stark socioeconomic differences between rural (mostly in apartheid homelands) and urban areas – has strong historical and political roots, and is another reason why it has been pushed up the policy agenda. There is emphasis on addressing both rural development (National Planning Commission, 2013) and to provide city support (City Support Programme, 2022). Political debates are also, at times, centred around local economic conditions. For instance, political parties use claims about rapid local population growth (Sacks, 2014) and "low" relative local unemployment (Pretorius and Clifford, 2019) to support their policy messaging. These claims are, at times, contentious, and it is therefore essential to base them in reliable data.

Data is therefore important for evidence-based policy, but also to more accurately and objectively frame social and political narratives. Building on previous work (Coetzee, 2012; Bezuidenhout et al., 2021), this paper surveys the data landscape for monitoring metropolitan economies in South Africa, and considers whether traditional or less conventional sources are required to provide an adequate view of urban agglomerations, evidence to support urban planning and to understand the place of South African cities in the country's development. However, this paper makes a more direct attempt to establish which of the available sources provide the most credible information. The hope is that this paper serves as a catalyst for an ongoing process by which new, value-adding indicators are brought into the debate, but also a continuous process of interrogating the status quo of the evidence, and whether indicators are suited for purpose.

2. Spatial inequality and the rise of the metropole

Apart from bureaucrats and politicians, academics pay significant attention to regional aspects of the economy. This is because spatial inequality is on the rise, especially in developing countries. Efforts to measure local economic activity have gained prominence in the development literature. These regional disparities are rooted in historical factors, but have been reinforced by processes of human migration and urban industrial agglomeration (McKay and Perge, 2015; Kanbur and Venables, 2005). In particular, new economic geography theories emphasise the role of scale economies in creating cross-regional inequalities (Venables, 2016; Krugman, 1991): on the one hand, urban growth can be an engine for developing national economies; on the other hand, urban growth creates undesirable regional economic cleavages that can leave rural areas behind – especially if migration from rural to urban areas is selective. Spatial inequalities are usually cited to motivate for the implementation of place-based economic policies that support areas that have been neglected or to prioritise potential urban economic hubs. However, despite the salient impacts of and social imperative of addressing spatial inequality, there is no consensus that place-based policies are theoretically defensible or effective in practice (Moretti, 2011). Be that as it may, spatial inequalities are hard to resolve and future eco- nomic development will have an increasingly metropolitan focus. Understanding aggregate economic growth therefore means – to a large extent – understanding metropolitan growth.

Because especially developing countries are rapidly shifting towards spatial equilibria characterised by an urban economic core, there is a high burden to collect new, timely and credible regional data in these data scarce contexts. Workers in developing countries that are further along the development trajectory tend to leave agricultural self-employment to enter wage-paying sectors in mainly urban areas

(Gindling and Newhouse, 2014). In South Africa, the abolition of apartheid -era influx controls in 1986 led to rapid urbanisation, especially through women's migration to join the labour market (von Fintel and Moses, 2017; Turok, 2012). Duranton (2015) and Chauvin et al. (2017) emphasise that urban agglomerations are as important in understanding developing economies as they are in the developed world. South Africa is not different. Spatial inequality remains defined by historical policies of separate development which created a long-run economic divergence that favoured urban areas and marginalised former homelands (Noble and Wright, 2013; von Fintel, 2018; von Fintel and Fourie, 2019).

3. Measuring Metropolitan Economies

Because cities are likely to dominate future growth in the developing world it is essential to find better ways to measure this part of the economy. The architectures of the first national statistical systems were primarily designed to track movements in aggregate national economic activity. However, even the pioneers of national accounting considered the distribution of economic activity to be an important component in giving effective interpretation to national aggregates (Kuznets et al., 1941). While this pertains to concerns with distribution more generally, inequalities across location lay an empirical floor for the extent of overall economic inequalities. In fact, Kuznets (1955) grounded his famous "Kuznets Curve" in arguing about spatial inequalities between urban and rural areas. However, comparable data that are representative of multiple areas and collected and post-processed to be harmonised, are rare to come by.

Traditionally economic measurement has relied on household and firm survey data, in addition to the use of administrative records. Using surveys to measure economic activity has grown in prominence, but significant doubts have been raised about the quality of this data. A long-standing concern that survey data and national accounts trace different growth trajectories raises questions about how countries and regional economies should be quantified. In particular, measurement has been complicated by quantifying unrecorded informal agriculture and other self-employment entrepreneurial activity in developing economies; it has been further complicated by statistical agencies' limited capacity and political sensitivities around collecting and releasing public data (Jerven, 2013; Devarajan, 2013; Deaton, 2016).

Before exploring the breadth and depth of sources for measuring metropolitan economic activities in South Africa, we briefly explore the uses and limitations of four broad types of data.

3.1 Household surveys

One of the most dominant sources of data comes from household surveys. From the labour market perspective, they measure economic activity on the supply side, and are suitable for gathering information on employment, population and remuneration. They also offer the benefit of observing individuals and households that are not engaged in economic activity – for instance, administrative tax records have no information on the unemployed.

Newer household surveys include detailed location information, especially when the sample is designed for generating small area statistics.[3] By the nature of the survey design, geographic information is linked to individuals' residential locations. Labour market statistics are therefore not necessarily representative of the locations where work was performed. This mismatch is only problematic if regions have typically long commuting distances, and if the research question demands that granular geographic differences within cities must be monitored. However, if metropolitan areas are analysed in totality, this is less problematic. In this case, daily movements between residential and work places are likely to be confined within the boundaries of cities. Detailed analysis within South African metropoles should, however, take a more careful approach that takes this mismatch into account, especially because distances between residential and workplaces are comparatively long in South Africa (Kerr, 2017).

Putting the definition of geography aside, there are other well-known problems with survey data that are more generally applicable, but which also affect analysis of specific regions (Deaton, 2016). Self-reporting errors are common (Ardington et al., 2006; Wittenberg, 2017a). Furthermore, data are not necessarily comparable over time, limiting the possibility to assess changes in local economic activity. Changes in survey

[3] For instance, GPS co-ordinates of the survey clusters are included in the World Bank's Living Standards Measurement Surveys, the Demographic and Health Surveys and Afrobarometer. When surveys are stratified to specific regions, they also contain regional indicators.

design and data post-processing procedures can compromise both national and regional comparisons of indicators over time (Kerr and Wittenberg, 2019). Furthermore, samples in household surveys are not sufficiently powered to analyse sub-municipal differences in outcomes. These issues are discussed in greater detail in relation to specific variables in later sections.

3.2 Administrative records

Administrative records – usually collected by tax authorities or other government departments – present a credible alternative to survey data and come with some distinct advantages. Firstly, because of legal mandates to report on all income and employment to tax authorities, sampling biases are alleviated and measurement errors from misreporting reduced. Secondly, administrative records are captured in similar ways over time, facilitating long-run comparisons. Thirdly, because firms have a role in reporting to tax authorities, it is possible to monitor the demand side of the labour market. That is, firm characteristics are known, including their total employment, firm location and their turnover from production. However, this type of data also has limitations. Firstly, the data contain no information on the unemployed, and firms and employees beyond the formal sector typically go unrecorded. Secondly, while firm locations are collected, it is not a given that the information is superior to the residential locations collected in household surveys. Because firms report the location of head offices and not for each branch or plant, economic statistics of larger companies with wide spatial coverage may be incorrectly allocated to main urban centres. This "headquarter bias" is likely to overstate production and employment in economic centres with strong agglomeration economies, and indicators will be understated in areas with many secondary operations (Nell and Visagie, 2022). These patterns imply upward biases in estimates of spatial inequalities between urban and rural areas; even within the urban sector, larger metropolitan areas could appear more productive and employment-intensive than expected, especially if headquarters are concentrated mainly in the largest metropolitan areas.

3.3 Unconventional sources

Household and firm-level data are usually collected at discrete points in time and do not always cover the full population of interest. The age of big data provides real opportunities to use proximate measures that can be tracked almost in real time, and can cover most of the globe. Luminosity from night lights is a prime example of this approach and has become a core measure in the economists' toolkit, as famously shown by Henderson et al. (2012). It stands as a close proxy for Gross Domestic Product, has global coverage and is recorded at high time frequency – importantly it gives information on areas that are not typically observed in surveys. Furthermore, the data are freely available and do not require the same human effort that statistical agencies must put in to administer large scale surveys. However, the data only measure economic variables indirectly, and must be modelled to convert them into interpretable monetary terms. It is, for instance, not possible to assess which people and sectors were responsible for the production and/or employment represented by brighter lights. However, in the absence of other local economic indicators, economists have increasingly relied on sources like these to measure regional economic growth (von Fintel and Moses, 2017; Obikili, 2019; Gibson et al., 2021).

Jean et al. (2016) illustrate how combining satellite and traditional survey data can also improve poverty measurement across geographic space. Remotely sensed data monitoring daylight landscapes have also un- locked the potential to monitor social change in response to policy and environmental shocks at a regional level. This approach has been particularly useful in rural settings to monitor both recent and historical impacts of events on agricultural activity (Fouotsa Manfouo and Von Fintel, 2022; Chingozha and von Fintel, 2019). Most recently, however, daytime satellite data have outperformed night lights data in predicting in- come and population for US small areas (Khachiyan et al., 2022). This work highlights that daytime satellite data – though more difficult to use than night time lights data – will likely become an important input for the analysis of urban spaces. The South African Cities Network (Mudau, 2019) uses night lights to verify the outputs of their demographic population models; they also emphasise the use of geotagged internet activity and social media use, as well as cellphone activity to monitor the movements of people. However, night lights data were more suitable in this regard.

Glaeser et al. (2018) interrogated other new ways of generating urban economic indicators with unconventional "big data" such as Google Street View images and internet-generated data from search engines, social media and the private sector. Combining Google's imagery with machine learning and citizen science (which leverages the power of the crowd) to rapidly differentiate among economically important objects in urban landscapes, such as (in)formal housing and businesses (Chingozha and von Fintel, 2018; Rebelo, 2020).[4] The scale of these applications currently remains small, but with increased resolution and computing power, the possibilities for measuring economic activity at high time frequency and with geographic granularity is within reach. This increases capacity to identify shocks in timely ways and also allows observers to pinpoint shocks to specific locations.

4. Inventory of main sources to monitor South Africa's metropolitan economies

Tracking sub-national economic activity is not a new endeavour. Coetzee (2012) took stock of available data for monitoring regional economic activity in South Africa in 2012. He lamented the limitations of existing sources, as implied by the quotation marks around the word "available" in the title of his paper. This exercise was published at the time when the Western Cape Government introduced its Municipal Economic Review and Outlook (MERO), which – up until 2022 – continues to rely on data from Quantec's Regional Service (Western Cape Government, 2022).[5] An updated inventory by Bezuidenhout et al. (2021) showed that the situation around data availability had not changed significantly. There are a handful of exceptions. While Coetzee (2012) anticipated the potential benefit of using administrative data from the South African Revenue Sources, Bezuidenhout et al.'s (2021) update confirmed the recent availability of the spatial tax panel dataset produced by Nell and Visagie (2022). Another difference between the first and second assessments is a list of potential new strategies for

[4] Rebelo (2020) used remotely sensed data and machine learning (https://alannarebelo.users.earthengine.app/view/ housingapp) to show how the spread of housing increased in a part of Cape Town. The tool illustrates how satellite data can be used to monitor real-time migration responses to the COVID-19 pandemic.

[5] The objective of the MERO is to "inform the Western Cape Government's evidence-based approach towards integrated planning and budgeting by guiding the equitable and sustainable distribution of financial resources in support of local economic development (LED) and growth. The MERO provides socio-economic intelligence at a municipal level, which feeds into municipal integrated development plans (IDPs), spatial development frameworks (SDFs), LED strategies, municipal reporting and the budget process of municipalities." (Western Cape Government, 2022)

unlocking sources of regional data, though most have not materialised to date.

While both papers considered the potential for using inaccessible firm-level data from the Companies and Intellectual Property Commission, the situation in this regard remains unchanged at the time of writing. Coetzee (2012) also wrote his review at the time when the use of satellite data was only emerging. Subsequently, this approach has been widely implemented in South Africa at the ward (Obikili, 2019; Matarutse, 2021), municipal (von Fintel and Moses, 2017) and provincial levels (Coetzee and Kleynhans, 2021) for answering a range of empirical questions.

The discussion reveals that regional economic data availability has been on a slow-moving trajectory, but the demand for such information has remained robust from academics, local and national governments and private companies. This paper will not repeat the recent stocktaking exercises on data availability in an exhaustive way (Coetzee, 2012; Bezuidenhout et al., 2021). Rather, the objectives of this paper are to firstly collate the most accessible sources of data, secondly to compare time trends across sources in all of South Africa's 8 metropolitan areas, and finally to assess which data source is most broadly credible across regions, and which can credibly replicate accepted empirical relationships.

The dimensions that will be explored are metropolitan population sizes (including the total population, the working-age population and population density), employment (including total employment, formal sector employment and sectoral composition), output (including Gross Value Added by sector and as proxied by the VIIRS luminosity data) and remuneration per worker (including all workers and formal sector workers).

4.1 Census and Survey Data

4.1.1 Census and Mid-year Population Estimates

The most important source of public data comes from censuses conducted by Statistics South Africa. They benchmark national population totals, and – importantly – also do so at the regional scale. However, censuses are infrequently collected and costly to implement. The latest information comes from the 2011 census, with the 2022 installation pending. In the interim, (local) population estimates are imputed based on supplementary demographic data and modeling assumptions. Statistics South Africa publishes annual mid-year population estimates using this approach. Importantly, these estimates are grounded in the "census gold standard" for up to a decade after it was enumerated. In the interim, population estimates are adjusted relying on assumptions about fertility, mortality and particularly international migration rates derived from supplementary statistics by location. Geographically disaggregated estimates additionally rely on incorporating rates of internal migration across regions to update local population totals. However, there is limited information in this regard in the inter-census years, and Statistics South Africa indicates that it awaits the outcomes of Census 2022 to update their assumptions about the rates at which people relocate (Statistics South Africa, 2022). When there are periods of rapid movement between locations4, the methodological status quo entails that comparisons across regions are less reliable. Importantly, citing changes in local population estimates to infer patterns of in-migration into cities during inter-census periods is not justified – such comparisons only recover the assumptions about migration introduced to the demographic models, and not the actual movement of people. Verified, up to date population information is therefore hard to come by. Satellite data that monitor settlement patterns could, in the future, become a more widely used method to obtain up-to-date information in inter-census years. Current solutions using satellite data, however, have limited geographic coverage (Rebelo, 2020) or are only published twice a decade (NASA, 2020).

4.1.2 Quarterly Labour Fore Survey

Despite the limitations of mid-year population estimates for monitoring population changes within regions, they nevertheless play a crucial role in developing sampling frames and weighting schemes for other surveys. In particular, the South African census forms the basis for designing its household surveys, another important input into monitoring labour market statistics in city regions. Sample surveys – such as the Quarterly Labour Force Survey (QLFS) or the General Household Survey (GHS) – start with a set of strata (regions, including metropolitan areas, or sub-populations for which the sample data are forced to be representative), within which geographic "clusters" are sampled. In turn, households are sampled within the clusters. The clusters are drawn from a set of representative enumerator areas from the census called the "master sample". The design of the survey ensures that clusters are more likely to be selected in proportion to their population size in the census. Such a design requires that survey weights be applied in all statistical computations to ensure that sample statistics are consistent with the composition and size of the population. In particular, households (and, by implication, regions) that are over-sampled are down-weighted, and conversely under-sampled households are upweighted. However, the weights are calibrated to mid-year population estimates. All local economic indicators derived from these data are therefore inherently dependent on the validity of regional population estimates. If population estimates fall out of line because of invalid assumptions on fertility, mortality and migration, so do household survey weights. Consequently, statistics produced from these sources may become less reliable the more time passes since the last census.

The QLFS strata are designed to be representative within each of South Africa's 9 provinces, and in particular within each of the metropolitan areas. When using appropriate weights, it is therefore possible to generate statistics over time that – as far as assumptions are correct – appropriately reflect the economic conditions of South African metropoles. However, surveys are not designed to draw reliable inferences about sub-metropolitan regions.

In addition to weights that can become outdated, sampling frames may poorly reflect the population in years long after the census on which they are based. Before the first quarter of 2015, the QLFS survey used a master sample based on the 2001 census, after which there was a transition to a new sampling frame based on the 2011 census. Kerr and Wittenberg (2019) discuss the implications of delaying updates to the sampling frame. At the time of the transition to the new sampling frame, South Africa's unemployment rate increased discontinuously. They argue that this shift occurred because the master sample drawn from the 2001 census may have excluded areas that had rapid population growth leading up to 2015. These growth points, many of which were located in Gauteng, did not create jobs as rapidly as the growing population and therefore had higher unemployment. Figure 1 confirms a spike in Johannesburg's labour force participation rate, that accounts for a sudden rise in unemployment in the city.

Once these enumerator areas rejoined the master sample, they introduced an upward correction to unemployment rates. Though a similar, temporary surge in labour force participation occurred in Ekurhuleni, this metro also experienced a temporary surge in employment rates, resulting in a stable unemployment rate. While this particular unemployment series seems to be consistent over time, its underlying components have changed with the sampling frame. In Nelson Man- dela Bay, by contrast, new enumerator areas had suddenly declining labour force participation – consistent with out-migration – combined with stable employment rates to produce temporarily lower unemployment statistics. These patterns emphasises the implications of not measuring the full extent of migration into or out of urban spaces for designing surveys that are used to generate urban economic indicators.

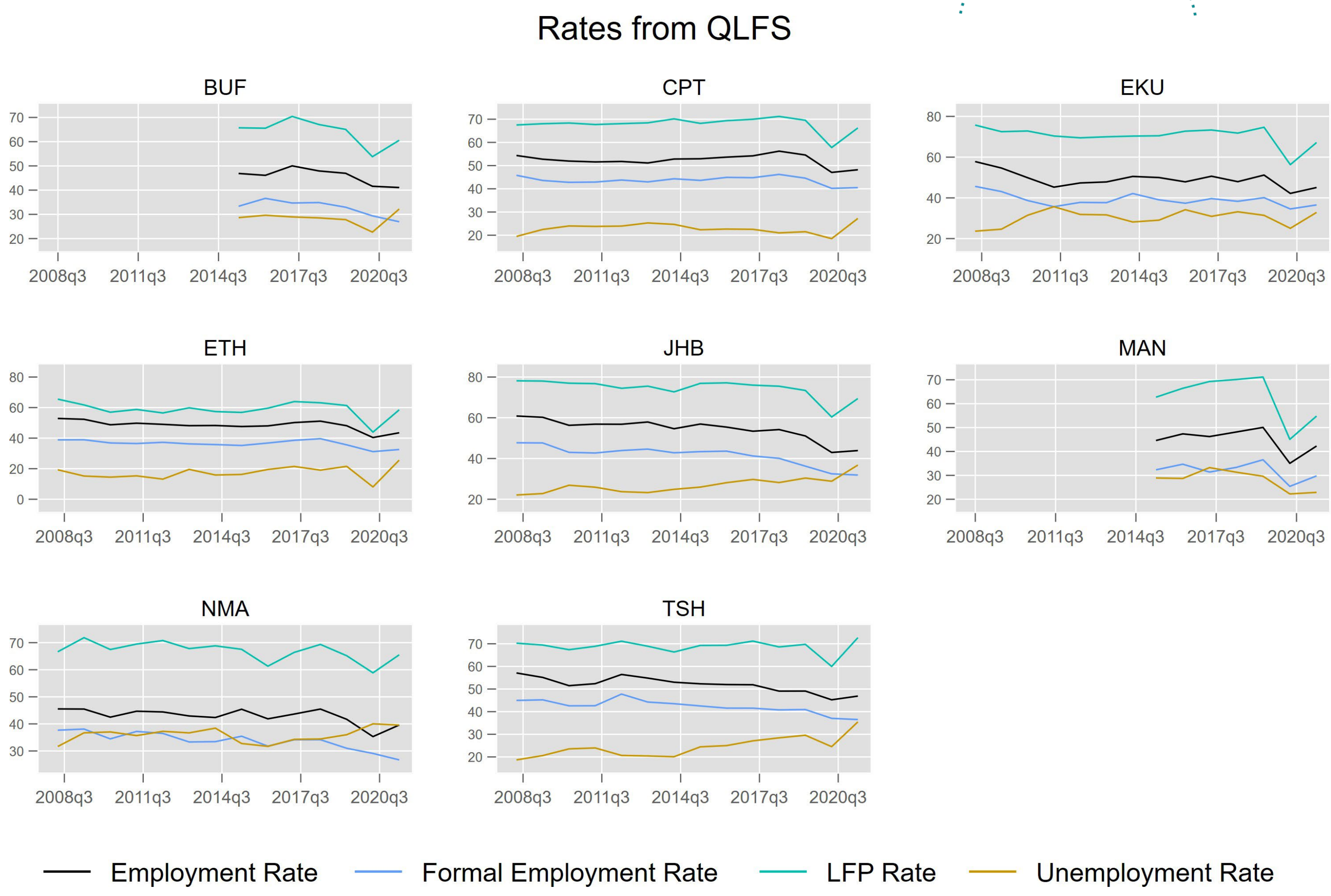
4.1.3 Post-Apartheid Labour Market Series

A related source is the Post-Apartheid Labour Market Series (PALMS), produced by DataFirst at the University of Cape Town (Kerr et al., 2019). It attempts to correct for some inconsistencies in QLFS survey design by harmonising and post-processing a range of original sources, including Statistics South Africa's October Household Surveys from 1993 to 1999, Labour Force Surveys from 2000 to 2007 and Quarterly Labour Force Surveys from 2008 onwards. In particular, the producers issue a set of alternative "cross entropy" weights that better reflect demographic changes forecast by the CARe actuarial model (Machemedze et al., 2020). They also ensure that variables are coded in the same way across time, and create earnings estimates with consistent definition across time. The post-processing improves comparability over time and makes the data friendly for use by researchers. However, the drawback is the lag with which the data are released - the current version extends only to 2019, which is almost 3 years behind the public release of QLFS data from StatsSA. The lag entails that any user that requires very recent information (such as policy makers), would be more inclined to use the original, unharmonised sources than the post-processed PALMS.

PALMS makes an important contribution by releasing – as far as is possible – a credible earnings series. Its construction and utility are documented by Wittenberg (2017b) and Wittenberg (2017c). A range of influential reporting issues are addressed by this process. Firstly, many respondents decline to respond to questions on their earnings, because they consider this a confidential piece of information. Wealthier individuals may wish to conceal the size of their incomes, whilst low-paid informal sector workers who have unstable income sources may find it difficult to quantify their exact earnings. In addition to non-response, many South African survey respondents are likely to report that they receive zero income, despite having characteristics that put them in a position to earn far beyond the poverty line (Ardington et al., 2006). This response behaviour is also consistent with respondents wishing to maintain the confidentiality of their income information. The result is that earnings are under-reported. Top earners report earnings in surveys that are about 40% below those recorded in administrative tax data (Wittenberg, 2017a). On the other hand, many outliers to the top are included in the data sources and are removed from the PALMS. Secondly, surveys are designed to allow respondents to report their incomes in brackets, including an unbounded top bracket for high earners (Fintel, 2007). This option increases item response rates on income questions – especially from high earners – but draws responses from a systematically different group compared to those who offer an exact point response.

PALMS provides two options for addressing these statistically meaningful problems for surveys spanning the period 1993 to 2019. Firstly, it provides so-called "bracket weights", allowing users to limit their sample to respondents who reported exact earnings amounts, and reweighting their earnings to reflect the distribution of the broader population. Secondly, a set of sequential multiple regression imputations, accounting as far as possible for all forms of mis- or underreporting, are released alongside the main files. However, there are limitations to this approach. Wittenberg (2017b) shows that distributional statistics can only be reliably estimated until the second quarter of 2012. As Kerr and Wittenberg (2019) highlight, using data beyond that period produces erratic and implausible distributional statistics. The reason lies in the fact that Statistics South Africa produces its own imputations to account for bracket and non-response, and in recent years they do so without releasing the original data containing missing values, outliers, zeroes and brackets. It is therefore not possible to implement a full range of independent and consistent corrections to harmonise the

Figure 1: Labour market rates by metropolitan area: Various sources



data. Importantly, further investigation reveals that the Statistics South Africa methodology for imputing data changed in 2012, but the process was not documented. This break in the series highlights not only the importance of measuring statistics in a consistent manner over time, but also following consistent post- processing methodologies that are transparently communicated.

4.1.4 Alternatives

These findings emphasise that measuring changes in incomes over time – especially more recently – depends on using sources other than the QLFS. By natural extension, this also has bearing on estimating changes in metropolitan earnings in future. Kerr and Wittenberg (2019) propose that the SARS data has filled a gap since it has become available, though it comes with its own limitations, as discussed in section 4.3.2. However, there are other potential sources. The annual General Household Survey (GHS) has been available since 2002, and comes with unimputed earnings data that are compatible with the PALMS methodology (Wittenberg, 2017a). Kerr and Wittenberg (2019) show that inequality estimates closely track those from the (Q)LFS before 2010, but beyond that the estimates diverge. This gives further support to the suspicion that QLFS data are unreliably imputed by Statistics South Africa. Apart from problems arising from a static master sample, the continuity of this source may prove a useful tool to monitor changes in incomes in future. Another prominent source of income data comes from the National Income Dynamics Study (NIDS) (Southern Africa Labour and Development Research Unit, 2017). Its main benefit is that respondents were asked to provide comprehensive information on all sources of incomes, and not only labour market earnings. However, there is significant evidence to show that the quality of earnings data is compromised by significant measurement errors (Burger et al., 2016) – a problem that is not unique to the NIDS, nor to South Africa, for that matter (Deaton, 2016). Furthermore, the 2008 NIDS data were stratified at the district council level, so that data were representative of metropolitan areas at that time. However, migration in and out of cities does not guarantee that the longitudinal sample remained representative of the original geography in subsequent waves of the data. While the metropolitan samples remain large, there are no guarantees that they offer a true reflection of these areas. Most pressing, however, is that the last round of this survey took place more than 5 years ago in 2017, and that it is unlikely to remain a useful source of data for tracking metropoles into the future.

Finally, Statistics South Africa's Quarterly Employment Survey (QES) is a potential additional source of survey data, but is unlikely to become publicly available (Kerr and Wittenberg, 2019).

This review of available survey data reveals that there are a number of limitations in measuring local economic indicators with these sources. Prominently, tracking local incomes is the most challenging aspect in this context.

4.2 Data generated outside of government agencies (Private Sector, Civic Sector and Academic)

4.2.1 Quantec Regional Service from Easydata's Subscription Service

Local governments and researchers rely heavily on Quantec's Regional Service (https://www.easydata.co.za/service/regional-service/) for local economic statistics. A range of indicators are reported annually for sub-national demarcations, including aggregates for metropolitan areas and sub-municipal electoral wards. The series' are standardised and easily accessible to subscribers. The source is unique in that it collates the widest range of regional economic data in the country over the longest period of time[6] and is the most user-friendly source.

A subset of variables is analysed in this paper. Firstly, total metropolitan population numbers and population density are extracted. While Quantec also provides the StatsSA mid-year population estimates for download, their own series is not identical to the original source. Judging by their high correlation, the Quantec series is presumably derived from the Statistics South Africa estimates. Secondly, Quantec provides statistics on total employment and formal employment, and can be disaggregated to SIC7 1-digit classifications. The third series is total annual

[6] The period covered depends on the variable analysed.

metropolitan remuneration, deflated to millions of 2015 Rands. A similar figure is given for formal sector remuneration. Both series' are converted to monthly figures and divided by relevant employment numbers to arrive at average remuneration per worker.[7] Finally, Quantec is the only source to provide output statistics at metropolitan level at annual frequency. Statistics South Africa, until recently, published Gross Domestic Product figures for provinces, while Quantec offers Gross Value Added numbers for metropoles, deflated to millions of 2015 Rands. The other dominant source of geographically disaggregated economic activity comes from Council for Scientific and Industrial Research (2018). Their "mesozone" data is widely used, but has last been updated in 2018. Presumably this source serves as an input for producing the Quantec data, though this link cannot be verified.

Quantec's data are modelled from other primary sources, including surveys from Statistics South Africa. However, the methods used to arrive at the figures are not publicly documented so that the process of moving from primary to final data is not reproducible. Many of the original sources used to generate the aggregates are only available at discrete points in time. In other instances, the original sources are only representative of demarcations larger than the sub-national/sub-municipal units reported by Quantec. By inference, the producers must rely on temporal interpolation and a modelling strategy that disaggregates the data to lower geographic units. It is unclear what empirical processes are followed, and Bezuidenhout et al. (2021) conclude that "this lack of transparency can be a constraint".

Nonetheless, the source has significant traction. Because local governments extensively use this service (Western Cape Government, 2022), it has implicit endorsement from the public sector. Researchers also appear to endorse the service. Figure 2 shows the number of documents associated with the keywords "Quantec Regional Service South Africa" released over time on Google Scholar and extracted using the "Publish or Perish" software (Harzing, 2007).[8] There has been a steady

[7] The figures are also inflated to 2016 Rands to match the PALMS earnings data.

[8] Documents include academic journal articles, policy reports and working papers that are listed on Google Scholar. A total of 962 papers were found, 69 of them had no date. No checks were done to verify that each paper actually used Quantec's Regional Service, as opposed to only referring to the source.

increase in the number of documents referring to or using Quantec's Regional Service, from close to none in 2005 to around 80 in recent years. Figure 2 also shows the average number of citations per year that each document receives. Publications attract about 1 to 2 annual citations in most years, though documents in some periods have higher counts.

The community that uses this data is therefore active. A similar search was done using the Scopus database. This database only indexes accredited, peer-reviewed journals. In general, paper and citation counts on this search engine are lower than on Google Scholar and signals a higher bar of peer review. Not a single publication could be found using the same search terms on Scopus. This suggests that while the Quantec Regional Service is widely adopted by the authors of a range of documents, the research is generally published in lower ranking journals or in outlets that are not peer reviewed.

4.2.2 Civic Sector: South African Cities Network

Free to use public data are also generated by civic sector networks that operate outside of academia, the private sector and government. For instance, the South African Cities Network offers one of the few demographic models of population for small areas within metropoles (Mudau, 2019). They emphasise the need for granular sub-metropolitan data because "electoral wards are at the heart of budget allocations, development initiatives and revenue generation". Their projections depart from standard modelling assumptions about migration, fertility and mortality. This is because migration statistics are not available at the ward level. They therefore follow simple parametric approaches by fitting quadratic curves through 1996, 2001 and 2011 census and 2016 Community Survey metropolitan population totals to interpolate populations in inter-census years and to predict populations into the future. They disaggregate their projections to ward-level and verify trajectories with night lights luminosity and digital connectivity data. Comparisons between these projections and those from other sources show whether simple parametric assumptions are sufficient to capture more complex demographic assumptions and processes.

4.2.3 Academic Sector: Thembisa Model

The Thembisa model, produced by actuaries and epidemiologists at the University of Cape Town, has become the baseline population projection model for South Africa (Johnson and Dorrington, 2022). It was primarily developed with the view of updating previous models to incorporate the impacts of the HIV/AIDS pandemic and the rollout of Anti-Retrovirals on demographic trends. UNAIDS has adopted the outputs as their official statistics for South Africa. The CARe model builds on the Thembisa model, but allows for disaggregation by sex and race (Machemedze et al., 2020). While both models produce national and provincial statistics, they are not designed to track metropolitan populations.

Nevertheless, as discussed in section 4.1.3, the model has become an important input into recalibrating weights of public survey data from which metropolitan geographic estimates can be gleaned (Machemedze et al., 2020). With the exception of race-specific estimates, reweighted statistics that reflect CARe model totals are not significantly different from those based on official population estimates from Statistics South Africa.

4.3 Administrative Data

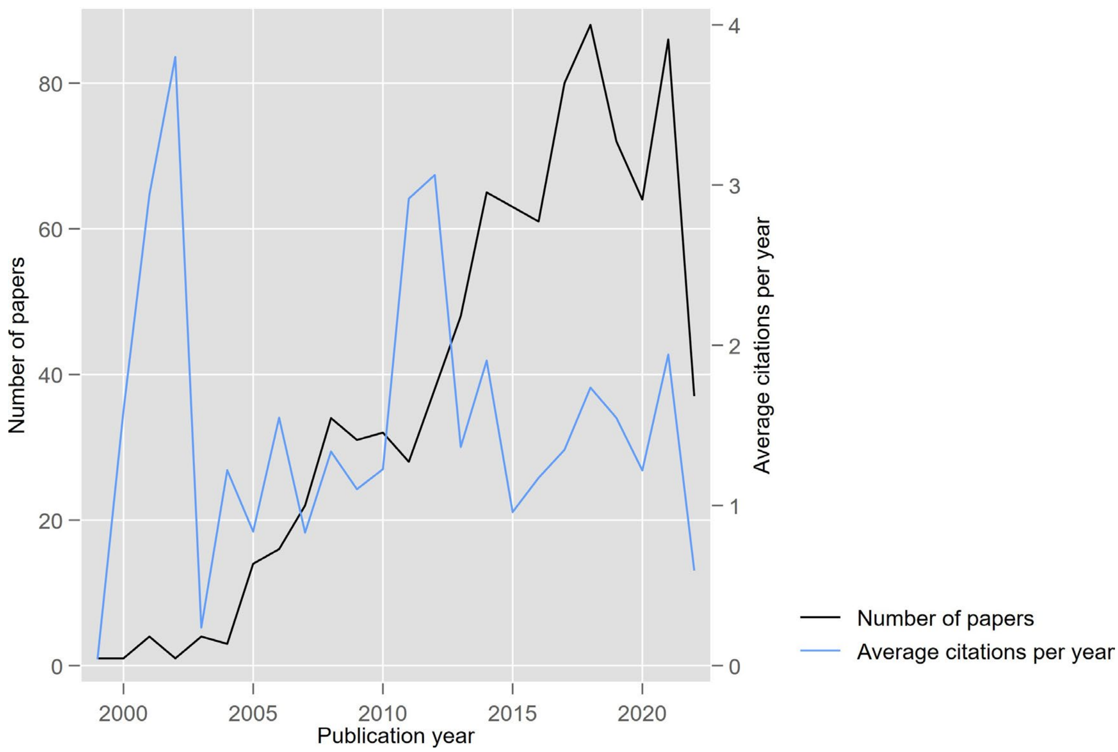
4.3.1 Unemployment Insurance Fund

Administrative records are relatively scarce in the South African data landscape. Kerr and Wittenberg (2019) raise the possibilities that could be derived from analysing data held by the Unemployment Insurance Fund. Employers are mandated to pay a proportion of their payroll to the Fund, which workers can reclaim when they are separated from their jobs. The implication is that the fund has almost exhaustive data on total (formal) employment, earnings and firm characteristics that could be leveraged for analysing (local) economic developments. However, access to these data for research and monitoring is limited.

4.3.2 South African Revenue Service tax data

The newest source of regional administrative data comes from the South African Revenue Service's administrative tax records for the tax years 2014 to 2021 (Nell and Visagie, 2022). The underlying records come from tax returns submitted by employers on behalf of employees (IRP5 certificates). A number of variables, including employment numbers and worker remuneration are aggregated to Uber H3 hexagons.[9] However, for the purposes of this paper, metropolitan statistics are analysed. These aggregate indicators are less noisy than sub-municipal indicators that rely on aerial weighting across postal codes.

Figure 2: Number of Google Scholar papers and citations using Quantec Regional Service



(Source Harzing (2007)) search conducted on 14 November 2022.

The individual- and firm-level data are only available to researchers in a secure laboratory with special permission from the National Treasury. However, the spatial data can be accessed remotely with permission from the Human Sciences Research Council. The process of producing the dataset is not yet fully documented, though introductory documents with metadata and some salient warnings accompany each new release. The most important of these is that the data are refreshed with each new

[9] The postal codes of firms link workers and employers to a particular location. Postal codes do not uniquely fit into the hexagonal demarcations, but the data are appropriately distributed across hexagons using aerial weighting. Hexagons with fewer than 10 employees are masked out to maintain confidentiality.

release of the firm-level data because firms may revise figures when they resubmit their tax returns. Additionally, this dataset suffers from "headquarter bias", as discussed in section 3.2. Importantly, the dataset best represents the formal part of the economy that constitutes the tax base.[10] The current analysis relies on the second version of the 2014-2021 data. While it is possible to analyse total firm revenues with the disaggregated data in the National Treasury's secure data lab, these variables have not yet been adopted in the latest spatial panel dataset. The current analysis is therefore limited to worker characteristics such as employment and remuneration.

Because IRP5 tax forms request employers to report on workers' employment status on a monthly basis, it is possible to construct quarterly aggregates for metropolitan areas, aligning the time frequency with that of the Quarterly Labour Force Survey. Industry-level aggregates (according to the 1-digit level of the Standard Industrial Classification version 7) are also available for metropolitan areas over time. It is, however, not possible to distinguish informal from formal work in this dataset – as noted above, it is assumed that the SARS data only adequately represents formal sector workers.

Two types of real monthly earnings variables are released with the data. Firstly, annual earnings of the median worker in each region is reported, and inflated to 2016 prices. This variable is the most straight forward to work with, but is not comparable to the mean remuneration per worker derived from Quantec data. Because of the well-known skewness inherent to earnings distributions, mean earnings are typically higher than median earnings; furthermore, time trajectories in earnings differ markedly across the distribution (Wittenberg, 2017b). This indicator can therefore not be directly compared to other sources (such as Quantec), which only allows for the analysis of average earnings. A second set of earnings information is therefore primarily used to estimate the mean over time and space. The SARS data reports the number of individuals who fall in a set of wage bands. An interval regression of bracket earnings (Fintel, 2007) on a set of time and metropolitan interactions, and weighted by the

[10] Reporting employers are only required to issue IRP5 certificates for individuals who earn more than R2000 per annum (Kerr and Wittenberg, 2019).

number of workers in each earnings bracket, gives relevant estimates of average earnings per (formal) worker in each metropole over time.

4.4 Night lights data

As noted in section 3.3, night lights data are the simplest source of satellite data to leverage for real time monitoring of local economies. The Visible Infrared Imaging Radiometer Suite (VIIRS) night lights data have been collected since 2011, both at higher geographic and time frequency than the DMSP-OLS series that it replaced. The data suffers from top-coding, because satellites are unable to distinguish the locations of large concentrations of light in highly urbanised areas. Fortunately, for the purposes of this article, VIIRS night lights products are more effective at tracking changes in urban economic activity than previous satellites were (Gibson et al., 2021). Matarutse (2021) presents a case study that illustrates the benefit of using this type of data. He combines ward-level luminosity data with city-level GDP from IHS Markit to produce a longitudinal GDP series for wards in eThekwini. While researchers are increasingly using this source for lack of alternatives, it is not immediately clear what type of economic activity is being measured, and how reliable modelling assumptions are to convert luminosity values into economic monetary terms.

5 Comparing metropolitan indicators in South Africa

A subset of the sources presented in 4 are compared. Each of South Africa's eight metropolitan areas is analysed separately to assess whether measurement biases arise only in specific situations. For instance, are statistics more reliable for larger cities with better information?

Figure 3 shows the location of the 8 metropolitan areas together with South Africa's provincial boundaries. Growth rates of VIIRS night lights data are plotted for 2019 (Figure 3a) and 2020 (figure 3b. Green dots indicate positive growth, while red dots indicate negative growth and yellow dots stability. Metropolitan areas are growth nodes, while negative growth is concentrated outside of metropolitan areas. However, the number of growth points declined during the COVID-19 pandemic.

5.1 Population Statistics

Population statistics are the point of departure for survey design and for constructing per capita measures of economic indicators. As noted before, they are also essential inputs for planning objectives and local budget allocations. The analysis therefore starts by discussing population trends. Figure A.1 illustrates population trajectories derived from various sources from 2008 to 2022. Statistics South Africa's mid-year population estimates – despite their dependence on a range of assumptions – serve as the benchmark. Apart from minor adjustments, Quantec appears to have adopted Statistics South Africa's figures as their own benchmark. Population estimates tend to grow gradually and smoothly at a near-constant rate each year in almost all metropolitan areas and across the full period of analysis. Population growth rates shown in figure A.2 are consistent with this observation. Except for Buffalo City and Mungaung, the estimated annual population growth rates using StatsSA data remain low and stable. This is mirrored in the linear growth of population density – the number of people per km2 – reported in figure A.3. These findings show that the demographic assumptions used to update population statistics from year to year remained mostly stagnant. Statistics South Africa started with census figures and mostly updated metropolitan population totals by small, predictable proportions. The close correspondence of the SACN series with those of StatsSA and Quantec suggests the simple parametric assumptions used by the former may be sufficient in most situations to capture the slow-changing demographic assumptions used by the latter. However, exceptional circumstances, such as the COVID-19 pandemic, break down the relationship. Reliable estimates crucially depend on up to date demographic assumptions or new census data.

The exceptions to most of these trends are found in Buffalo City and Mungaung, where population growth reversed. In Buffalo City, population growth – as per Quantec and Statistics South Africa – went as far as to decline. This changing trajectory arose almost discontinuously in 2012. Despite not recording an absolute decline in population, Mungang's trajectory followed a similar pattern to Buffalo city with slower growth. There are two possible reasons for this sudden change. Firstly, this could be the consequence of new information about in- and out-migration derived from the release of the 2011 census. Secondly, both cities attained metropolitan status in 2011, with boundary changes potentially affecting population numbers. However, in both cities there were more systematic long-run changes as shown in the StatsSA/Quantec figures. Population growth in Buffalo City was estimated to be negative in all periods after 2015 and the decline intensified from 2020 during the COVID-19 lockdown. Population growth slowed down around 2015 in Mungaung, with a rapid decline in growth that coincided with the COVID-19 lockdown period. While it is evident that the assumptions that generated these changes were linked across these two cities, it is not certain what those assumptions were – whether this was driven by assumptions about growing mortality, lower fertility or an acceleration in out-migration from these metropoles to other metropolitan areas. Apart from these specific cases, it is evident that demographic assumptions are slow to change, and that we may expect discontinuous adjustments when information from the 2022 census becomes available.

These exceptions also highlight the weaknesses of the SACN parametric projections, which were last updated in 2019 before the COVID-19 pandemic. Those projections initially follow similar patterns to the Statistics South Africa trajectory before the 2011 census. After 2011, the parametric assumption diverges from StatsSA's demographic assumptions which imply population declines in Buffalo City. Figure A.2 shows this most clearly. While SACN projections and StatsSA figures have similar growth rates in most cities, they are very different in Buffalo City and Mungaung. Model assumptions can therefore have a strong bearing on population projections, especially in smaller cities where in- and out-migration in response to social change and health/economic shocks may be volatile. It is therefore a difficult task to track population changes in these contexts without regular census information.

As expected, the total weighted population from the QLFS closely resembles the mid-year population – even the growth rates correspond well across sources from 2015. However, there are significant QLFS population fluctuations before 2015, and erratic year-on-year growth. Procedures for constructing weights were apparently updated around the same time that the master sample was updated in 2015. The weights seem better calibrated to reflect the stratification of the survey to the metropolitan level after 2015. It is not immediately clear how the procedure changed. Furthermore, in eThekwini, Nelson Mandela and Tshwane metropolitan areas, there were discontinuous jumps in QLFS population numbers around 2015, emphasising that procedures changed discontinously.

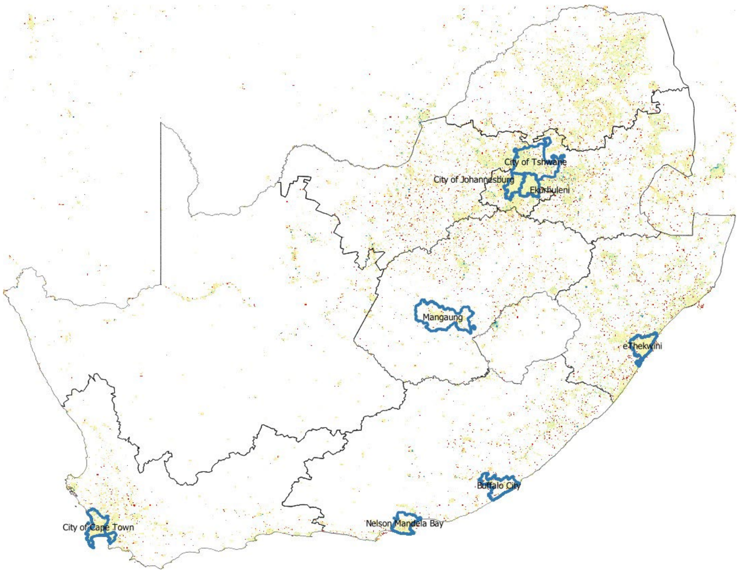
Figure A.1 also shows weighted QLFS population totals for the working aged population between 15 and 64, which is the basis for estimating labour market statistics. The levels are naturally lower than for the population as a whole, but time trajectories are mostly comparable to those of the whole population. Working-age totals using PALMS cross entropy weights are also compared. They match the QLFS estimates closely, highlighting that cross-entropy reweighting had no systematic relationship with metropolitan status. It is also consistent with Machemedze et al.'s (2020) findings that recalibrated weights only tend to change estimates for specific demographic groups.

5.2 Employment statistics

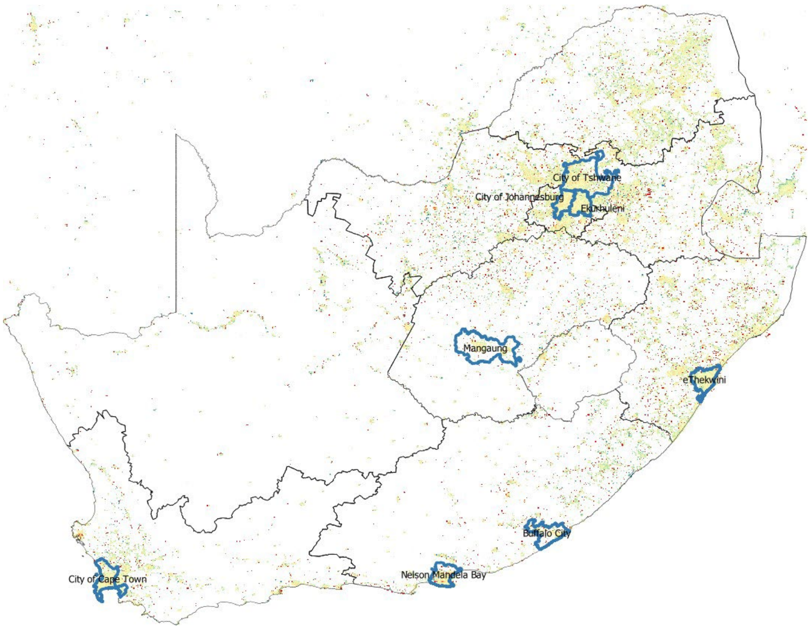
Figure B.1 compares total employment in each metropolitan area by source. Figures are also given for formal sector employment. Quantec figures are reported on an annual frequency. To match the "mid-year" benchmark for StatsSA's population statistics, other sources are limited to the second quarter of each year. Quantec figures align closely with those contained in the QLFS, signifying that they used the Statistics South Africa data as their primary source. It is furthermore apparent that Quantec smooths out fluctuations from its employment series. Some discrepancies are noticeable in the "smaller" metros of Buffalo City (where Quantec reports lower levels of formal and total employment than the QLFS), and Nelson Mandela (where Quantec figures appear to be underreported only before 2015).

In most cases, the SARS figures are closely comparable to the levels of employment reported for the formal sector in other sources. This affirms that the tax records are representative mainly of the formal economy. This pattern is especially apparent in some of the "larger" metros, namely Cape Town, eThekwini and Tshwane. This observation suggests that "headquarter bias" is absent or minimal in these particular "larger" cities.

Figure 3: Maps of metropolitan areas with Night Lights Growth Rates



a. 2019 VIIRS Growth Rate



2020 VIIRS Growth Rate

In Johannnesburg, however, employment trends from the SARS records correspond most closely to total employment levels and far exceed formal employment levels reported in other sources. It is unclear why this exception has arisen. On the one hand, it is tempting to conclude that figures are overstated because of "headquarter bias". That is, because many larger firms are registered in Johannesburg, employment attributable to their branches outside the city inflates this metro's figures. However, the uncannily close correspondence to the total employment figures casts some doubt on this explanation.

In "smaller" metros (Buffalo City, Ekurhuleni, Mangaung and Nelson Mandela), the SARS employment figures are much lower than those reported in the other sources, and the SARS time trend does not correspond to those in other sources in the same way it does in the larger metros. Again, there is a case to be made for "headquarter bias", but in the other direction. Firms that operate in these smaller metropolitan areas may have headquarters registered in the much larger cities. The work that is actually being done in the smaller metropoles is being recorded in the economic core of the country – according to the statistics this would account for the very high figures reported in the City of Johannesburg.

Overall, these findings suggest that Quantec has followed Statistics South Africa in constructing employment trends (but that the data have been smoothed), and that headquarter bias has different impacts on tax data associated with various metropolitan areas.

Figure B.2 shows that - apart from the earlier fluctuations contained in some of the data sources, annual employment growth rates derived from the range of sources correspond remarkably well with each other. This suggests that the various sample biases do not obscure time trends in the data, even if the levels are in dispute. This finding is encouraging, given that a similar pattern appears in data influenced by survey design bias (QLFS) and by unknown post-processing procedures (Quantec), and also in administrative data that is not primarily collected for the purpose of statistical analysis (SARS).

Do the data conform to other expected dynamics? If employment converges over time, the relationship between job growth (%Δemployment) and quarterly employment one year prior ($et-1$) should be negative. That is, times and regions with intially low employment levels catch up on other times and regions. Figure B.3 confirms that this relationship is insignificant (flat) or negative in most cases. Where it is negative, the strength of the relationship appears to be of similar magnitude in each of the sources.

This corroborates the notion that the various sources produce consistent dynamics and are reliable for understanding change.

However, despite the apparently consistent dynamics of these numbers across sources, the sectoral dynamics of formal employment point in another direction. Figures B.4 to B.6 decompose total formal employment into industry shares. All three sources reveal that the primary sector has negligible shares in total employment, as is to be expected in metropolitan areas. However, the SARS data rightly picks up on a noticeable (yet small) proportion of workers in mining in Ekurhuleni and Johannesburg, suggesting that this sector's presence in these cities is missed by survey sources that are not necessarily designed to enumerate rare events. On the other hand, a negligible share of employees in the SARS data work in the community and household sectors; by contrast, the QLFS measures a substantial proportion in the community sector, with no workers in the formal household sector; Quantec data shows that a large part of formal employment is located in the household sector. This latter finding is implausible because the household sector is mainly composed of informal work. The SARS data creates the impression that metropolitan employment is dominated by manufacturing and retail, while both sectors have far smaller shares in the other sources.

These findings suggest one of two problems with the sources. A first possibility is that workers in some sectors are missing from some of the sources – SARS data does not miss "rare sectors" such as urban workers in mining because of reporting mandates. But sample surveys are not necessarily designed to pick up "rare sectors". SARS data, however, misses workers in the household sector because of its formal sector bias. A second possibility is that all relevant workers are adequately represented in all the sources, but that their industries are being misclassified – either through self-report or in systematic ways related to survey design. Distinguishing between these possibilities requires further interrogation. If, however, Quantec uses QLFS as a benchmark for its employment statistics, it is not clear why they would classify a large segment of formal sector work in the household sector.

5.3 Earnings Statistics

As discussed in section 4.1, generating statistics on earnings is relatively complicated.

However, a range of indicators is available in each of the sources that were also used to analyse employment. As noted in section 4.1, there are various problems with QLFS earnings data. While Wittenberg (2017b) showed that post-processing could produce sensible and smooth earnings series at different parts of the distribution until 2012, this is not the case beyond that point. Figure C.1 shows that average monthly earnings from PALMS continued on an upward trajectory in some municipalities immediately after 2012, but that the series is erratic and unreliable thereafter. Outliers in selected years in some metropoles have significant influence on earnings statistics and trends. Figure C.2 confirms that growth rates in PALMS earnings per worker are unstable after 2012 and supports Kerr and Wittenberg's (2019) contention that other sources must be consulted in subsequent periods. SARS data could present as an alternative. SARS estimates do not correspond closely to the trends and levels reported in PALMS. However, in this article, these trends are estimated from bracketed region-level data and may have their own problems. It is therefore not certain whether the differences primarily lie with the survey data or unreliable methods applied to the bracketed administrative data. Indeed, figure C.3 shows that there are no systematic relationships between growth in mean earnings and lagged earnings as one would expect. It suggests that all sources have their limitations.

Where Quantec appears to have followed QLFS data to construct employment series, the findings from PALMS suggest why this approach is not plausible on the side of earnings. Quantec earnings estimates are generally higher than those from other sources, and follow an uncharacteristically smooth trajectory. They evidently used other sources and methods to construct earnings totals. The discussion above justifies organisations' efforts to explore alternative sources of earnings data. However, it is unclear what Quantec's source is. It is also not certain whether their earnings series' appear more stable than they really should be because of smoothing and modelling assumptions that are similar to the post-processing procedure used on employment data.

While various data producers and organisations have their own sources and post-processing procedures to create earnings series', there is no clear winner and no gold standard to guide users. Kerr et al. (2019) take a step in the right direction by applying the PALMS methodology consistently to the more flexible GHS series. However, it is essential that alternative sources with transparent methodologies should become available before wage differentials across metropoles and over time can be credibly analysed. The closest contender is the SARS data, from which better estimates may be produced. Figure C.4 shows that this data holds potential in this regard. The SARS panel also issues a nominal median wage series by metropolitan area, which is adjusted for inflation here. Remembering that median earnings usually fall below mean earnings, this series is not directly comparable to the trends presented in figures C.1, C.2 and C.3. However, the series' shown here are smooth and provide a realistic ranking of metropolitan areas, pointing to their possible credibility. Because median earnings in PALMS data stayed constant until 2012 (Wittenberg, 2017b), these additional findings lead us to believe that the subsequent period was the start of a declining trend. Figure C.3 also suggests that median wages started to diverge rather than converge. These factors require further investigation. Nonetheless, working with the teams that spatialise the SARS data could produce alternative comparaible and usable series' that pave a way forward in monitoring earnings in metropolitan areas in South Africa. However, SARS and many other sources, will not get around the omission of informal sector work from this type of analysis. Administrative datasets and formal sector companies are unlikely to contain records for these workers, and survey data remains the best barometer for this sector.

Other sources continue to remain outside of the public domain, but present as credible alternatives. The UIF data also covers a wide range of workers, but is also likely to omit informal work. Private sector payroll data may also be helpful, but with similar limitations to SARS and UIF data.

6 Outputs Statistics

The final set of statistics relate to economic output. Traditionally the smallest sub-national demarcation for which high time frequency GDP estimates have been produced has been provinces.[11] However, Statistics South Africa discontinued the release of provincial estimates in the first quarter of 2020, corresponding with the start of the COVID-19 pandemic in South Africa. Metropolitan output data are therefore scarce and out of date. However, Quantec releases a Gross Value Added series for fine geographic demarcations, including metropolitan areas. Again, it is not certain which methods were used to generate the small area series'. Given the lack of alternatives, these figures are compared with the trajectories generated by night lights luminosity data. Figure D.1 presents the comparisons of the trajectories. Broadly speaking, the night lights aggregate and Quantec GVA follow the same time trajectories across all metropolitan areas. In all cases, the metropolitan GVA figures point to a limited recovery after the initial shock brought on by COVID-19 lockdowns in 2020. However, this recovery is not detected by night lights data, the potential consequence of loadshedding that followed when the economy re-opened. Both GVA and night lights series exhibit initial growth, with turning points in all municipalities, consistent with the decline in real median wages noted in the SARS data (see figure C.4). Figure D.2 shows that Quantec and night lights produce the similar trends with small differences at various points in time. Apart from a few exceptions such as Johannesburg and Nelson Mandela Bay, the "convergence" relationships look similar between the two sources (see figure D.3).

Overall Quantec's metropolitan GVA data square up well with the VIIRS night lights. Some of the trends are also be supported by movements in the SARS wage data. On the one hand, this gives peace of mind that the Quantec estimates are consistent with other series. However, it is, of course, possible that Quantec used these sources as inputs to predict GVA, so that they are correlated by design. If this is the case, night lights indicators are not suited for verifying GVA figures – in fact, no indicators can then be used to validate the sub-national GVA data. It therefore remains uncertain whether there is a gold standard to speak of.

Nonetheless, metropolitan municipalities (Matarutse, 2021) and think tanks such as the South African Cities Network (Mudau, 2019) produce GDP and population

[11] The CSIR produces statistics for smaller "mesozones", but at low time frequency.

statistics for wards within cities. On the one hand, night lights have been used to verify population changes; however, geolocated internet and social media activity have been poorly correlated with forecasts (Mudau, 2019). Given that people move around cities significantly, they participate in the economy in a range of sub-metropolitan areas. Output in "their" ward is therefore perhaps an unnecessary concept, and city-level data may be a sufficient aggregate for planning purposes. This raises the question as to how useful it is to estimate statistics for very small sub-metropolitan areas, and even whether metropoles are the appropriate unit of analysis.

While Quantec produces GVA figures by sector (see Figure D.4), it is not possible to verify these against alternative data sources. Night lights, for instance, shine in the same way in all sectors and cannot distinguish betwen them. The composition of employment in other sources may offer one way to verify the plausibility of the sectoral shares found in the GVA data. The time trends and proportions are similar to Quantec's employment series, as evident in figures B.5 and D.4 – small differences may be attributed to GVA covering the whole economy, while employment was limited to only the formal sector. Again, it would be tempting to claim that the one source supports using the other. On the other hand, similar modelling assumptions may have been adopted in both sources, so they could be similarly flawed by design. One factor which potentially supports this composition is the declining share in manufacturing in both Quantec's GVA (figure B.5) and in QLFS formal employment (figure B.4). Yet it holds a stable share of employment in the SARS data (figure B.6), so that it is not unequivocal that Quantec's assumptions are correct.

An unexplored alternative is to use firm revenues from the SARS data to analyse regional output. This variable is available in the micro records in the secure lab, but has not yet been incorporated into the public release spatial panel. Given the potentially valuable information contained in the median wage series, the SARS data presents as a credible and well-documented alternative to other sources. The similar downward trend in real wages and night lights suggest that this is an avenue worth exploring. However, it is not certain whether headquarter bias will have more substantive influence on firm output than it does on employment. This is left as an open question for future research.

7. Discussion and conclusion

Researchers have made significant investments over many years to leverage existing sources and to unlock new data for analysing urban development in South Africa. The demand for data on cities robustly continues to exceed its supply. While the decade that passed between Coetzee's (2012) and Bezuidenhout et al.'s (2021) data inventories delivered a few new possibilities – the advent of the SARS administrative data and the growing use of satellite-generated data – the situation has remained relatively static. Cities have grown more rapidly than the information that is required to understand them better.

Using the prior inventories as a point of departure, this paper is the first to compare available sources with each other and to kickstart an ongoing process of verifying which indicators are most appropriate for tracking economic development in metropolitan areas. Both new and old data are analysed, including the "conventional" (sources such as government census data and household surveys), as well as the "unconventional" (administrative data and indicators derived from satellites). These sources are compared to widely-used data that are collated and processed for subscribers by a private sector provider, Quantec. The newer, unconventional sources expand the scope to interrogate the validity of some of the indicators that have been historically used. The analysis shows that while there are not yet agreeable "gold standards", the options for credible analysis are on the uptick. The way forward calls for uncovering primary data that is generally accessible, and for using transparent methodology with documented processes that are used to convert primary data to the estimated series used by researchers and stakeholders. Few of the currently available sources fully satisfy these criteria.

Localised population estimates remain at the core of urban planning and budgetary processes. Current modelling strategies depend on baseline census data and a range of modelling assumptions. Data derived from complex demographic assumptions and simpler parametric assumptions are equally (un)reliable in "normal" economic conditions. However, parametric models that do not adapt to rapid changes such as the COVID crisis perform poorly in these contexts; whether demographic assumptions have adapted appropriately is also uncertain. Least satisfying, however, is data for which the modelling assumptions are not known at all, as is the case for Quantec estimates. Importantly, users should be alert to the fact that studying changes in local population numbers using any series, only reverse engineers estimation assumptions, and does not necessarily reflect real demographic changes. The bottom line is that the more time has lapsed since a census, the less is known about local populations. It is likely that when Census 2022 figures are released, that existing assumptions about population growth will be challenged by new empirical evidence emerging after a decade of profound social, economic and spatial change. Because of these challenges, it is essential that data providers leverage the benefits of other sources that provide detailed population estimates in intercensus years. These include administrative data, internet generated data (Glaeser et al., 2018) and daytime satellite data (Khachiyan et al., 2022; Rebelo, 2020; Chingozha and von Fintel, 2018; NASA, 2020) to generate up-to-date indicators of settlement patterns. The domain of satellite data has expanded most rapidly in the last decade. It reduces reliance on irregularly enumerated census data and the need for complex demographic assumptions to update prior population estimates. However, there is still a way to go to produce population estimates for regions covering entire countries, and to do so regularly enough to track population movements close enough to their occurrence to have timely policy relevance. Another important consideration is standardising conventions for converting input data – such as physical structures identified by satellites – to population figures or economic indicators in monetary terms. Once these issues have been distilled, these possibilities show significant promise to provide regular updates on varying conditions between metropolitan areas and also in specific suburbs of these cities.

Employment is another outcome that is both practically important – local labour markets with many jobs on offer attract migrants and must plan for new public services – and also plays a part in political debates between rival local governments. While the sources analysed do not provide a consensus view of metropolitan employment levels, their annual growth rates are surprisingly robust across each of the sources studied. This consistency emerges despite varying surveying biases, weaknesses of sampling frames and weighting schemes, headquarter biases in firm data and non-uniform post-processing biases in all sources. At this stage, employment statistics appear to be the most reliably estimated among the set of metropolitan indicators in South Africa. Administrative tax data have significantly contributed to verify the size of the formal sector and has the unique ability to record employment in “rare sectors”, where surveys do not have the statistical power to do so. Two issues warrant further discussion. Firstly, administrative data from other sources, such as the UIF or the CIPC, could bolster current efforts and improve knowledge about local labour markets significantly. Secondly, however, a word of warning is necessary. Significant differences in the sectoral composition across the sources mean that the SARS administrative data cannot yet be elevated to a “gold standard”. While many stakeholders have access to sources such as Quantec or the QLFS, they should also be used with necessary caution. Especially the lack of transparency about post-processing of the widely adopted Quantec series means that the search for a better standard should continue.

The least reliable indicators relate to local earnings and economic output. Measuring earnings with surveys is notoriously difficult. However, working with and disclosing appropriate methodologies can mitigate some of these concerns. Quantec and Statistics South Africa surveys are both compromised by non-transparent post-processing methods, and for not availing the primary inputs used to derive earnings statistics. The spatialised SARS panel data shows promise in providing better metropolitan earnings series. Currently only median metropolitan earnings are available, but other statistics such as mean earnings per worker could be released for comparative purposes.

Since 2020, Statistics South Africa no longer releases sub-national GDP statistics. Quantec is the most widely used source for this purpose. However, their Gross Value Added figures can be broken down to sub-municipal level by using unclear methodologies. Given the strong correlations observed, it remains a possibility that local indicators were predicted from night lights data and other data inputs. A gold standard also remains elusive on this front.

7.1 Final observations

Data and local development have the potential to be symbiotic. “Good” statistics are essential inputs to formulating and monitoring policies that aim for improved economic development. However, the availability of harmonised, comparable statistics that follow transparent methodologies is – in turn – strongly dependent on the resources that come with economic development and are more widely available where administrations gather data as part of their management process. On a cross-national scale, high quality statistics are generated where countries have sufficient resources to build statistical capacity, a factor that has been lagging in developing economies (Devarajan, 2013). Efforts to deepen statistical capacity and improve data quality may also vary significantly across regions in specific countries. On the one hand, larger, more complex local economies are better equipped to gather and analyse administrative data. However, these local administrations also need this data to effectively manage local affairs. On the other hand, smaller, less developed cities and regions may not have adequate data gathering systems in place, but nevertheless have a need for high quality data to plan their way into greater regional prosperity. To the extent that data can help to formulate decisions that promote economic development, countries, regions and cities should prioritise resources towards creating local statistical capacity. In South Africa, larger city regions are responding to this priority. Gauteng City-Region Observatory (2008) has initiated cross-sectoral partnerships for generating local data. Cape Town’s open data portal is one of the first and largest of its kind, with some demonstrated local developmental impacts (Hlabano and Van Belle, 2019).

The situation is different in smaller cities and mainly rural municipalities. Solutions often rely on external capacity. Subscription services that standardise indicators suited to users’ needs fill an important gap in this market. However, this paper challenges users and providers to be critical about the data they rely on, and the statistical foundations and assumptions they are built on. While bureaucrats in poorer regions are unlikely to specialise to the degree that they will leverage challenging resources such as administrative and satellite data, there is nevertheless a call to academics, private providers and civic networks to offer this service, and to do so with reliable data inputs using transparent methods.

This paper has shown that while metropolitan data availability has been poor in South Africa for a long time, there are signs that the situation has started to improve – especially with the release of the SARS spatial panel in 2022. However, as data availability expands, other aspects of producing high quality data should come into focus. Firstly, publishing transparent methodologies – including descriptions about how data are collected, interpolated and post-processed – can improve the credibility of existing sources (such as the earnings data from StatsSA and the harmonised data from Quantec). Secondly and finally, the research community is tasked with establishing benchmarks for which series can be used as "gold standards", given the many weaknesses across the whole range of sources. To simultaneously achieve data availability, transparency and credible benchmarks will require co-operation between local and national government agencies with the research community and private organisations. As illustrated by the release of the SARS spatial panel, releasing existing administrative data (such as the UIF and CIPC data) can make a significant contribution to producing valuable information if public organisations and researchers can work together on these data. Further, data in the private domain (such as mobile phone location records and company data) can make a similar contribution to understanding metropolitan economies. Finally, researchers with the skills to process unconventional sources (such as satellite data), can play a meaningful role to synthesise the various data into actionable insights.

References

Ardington, C., Lam, D., Leibbrandt, M., and Welch, M. (2006). The sensitivity to key data imputations of recent estimates of income poverty and inequality in South Africa. Economic Modelling, 23(5):822–835.

Bezuidenhout, C., Malindi, K., and Rankin, N. (2021). A report and strategy for spatialised economic activity data. Technical report, City Support Programme, National Treasury, Pretoria.

Burger, R., Klasen, S., and Zoch, A. (2016). Estimating income mobility when income is measured with error: the case of South Africa.

Chauvin, J. P., Glaeser, E., Ma, Y., and Tobio, K. (2017). What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. Journal of Urban Economics, 98:17–49.

Chingozha, T. and von Fintel, D. (2018). Applying astronomy tools in the field of development economics. Proceedings of the International Astronomical Union, 14:589–590. Type: Article.

Chingozha, T. and von Fintel, D. (2019). The Complementarity Between Property Rights and Market Access for Crop Cultivation in Southern Rhodesia: Evidence from Historical Satellite Data. Economic History of Developing Regions, 34(2):132–155. Type: Article.

City Support Programme (2018). Managing Urbanisation to Achieve Inclusive Growth: A review of trends in South African urbanisation and suggestions for improved management of urbanisation.

City Support Programme (2022). CSP - Home.

Coetzee, C. (2012). Data resources 'available' for local economic assessment in South Africa. Skills at Work: Theory and Practice Journal, 5(1):1–12. Publisher: University of KwaZulu-Natal.

Coetzee, C. E. and Kleynhans, E. P. J. (2021). Estimating economic activity of South African provinces using satellite technology: An econometric panel data application. Remote Sensing Applications: Society and Environment, 23:100564.

Council for Scientific and Industrial Research (2018). stepSA - Spatial Temporal Evidence for Planning in South Africa.

Deaton, A. (2016). Measuring and Understanding Behavior, Welfare, and Poverty. American Economic Review, 106(6):1221–1243.

Devarajan, S. (2013). Africa's Statistical Tragedy. Review of Income and Wealth, 59(S1): S9–S15._eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/roiw.12013 .

Duranton, G. (2015). Growing through Cities in Developing Countries. World Bank Research Observer, 30(1):39–73. Accepted: 2016-08-04T19:22:44Z Place: Washington, DC Publisher: World Bank.

Fintel, D. v. (2007). Dealing with earnings bracket responses in household surveys - How sharp are midpoint imputations? South African Journal of Economics, 75(2):293–312. Type: Article.

Fouotsa Manfouo, N. C. and Von Fintel, D. (2022). Investigating the effects of drought and lockdowns on smallholder and commercial agricultural production in KwaZulu-Natal using remotely sensed data. Heliyon, 8(11):e11637.

Gauteng City-Region Observatory (2008). About The GCRO | GCRO.

Gibson, J., Olivia, S., Boe-Gibson, G., and Li, C. (2021). Which night lights data should we use in economics, and where? Journal of Development Economics, 149:102602.

Gindling, T. H. and Newhouse, D. (2014). Self-Employment in the Developing World. World Development, 56:313–331.

Glaeser, E. L., Kominers, S. D., Luca, M., and Naik, N. (2018). Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life. Economic Inquiry, 56(1):114–137. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10. 1111/ecin.12364.

Harzing, A. (2007). Publish or Perish.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring Economic Growth from Outer Space.

American Economic Review, 102(2):994–1028.

Hlabano, M. and Van Belle, J.-P. (2019). Tracing the Impact of the City of Cape Town's Open Data Initiative on Communities and Development. In Nielsen, P. and Kimaro, H. C., editors, Information and Communication Technologies for Development. Strengthening Southern-Driven Cooperation as a Catalyst for ICT4D, IFIP Advances in Information and Communication Technology, pages 284–294,

Cham. Springer International Publishing.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., and Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. Science, 353(6301):790–794. Publisher: American Association for the Advancement of Science.

Jerven, M. (2013). Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It. Cornell University Press, Ithaca.

Johnson, L. and Dorrington, R. (2022). Thembisa version 4.5: A model for evaluating the impact of HIV/AIDS in South Africa. Technical report, University of Cape Town, Cape Town.

Kanbur, R. and Venables, A. J. (2005). Spatial Inequality and Development. In Kanbur, R. and Venables, A. J., editors, Spatial Inequality and Development, page 0. Oxford University Press.

Kerr, A. (2017). Tax(i)ing the Poor? Commuting Costs in South African Cities. South African Journal of Economics, 85(3):321–340. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/saje.12161.

Kerr, A., Lam, D., and Wittenberg, M. (2019). Post Apartheid Labour Market Series 1993-2019 V3.3. Type: dataset.

Kerr, A. and Wittenberg, M. (2019). Earnings and employment microdata in South Africa.

Khachiyan, A., Thomas, A., Zhou, H., Hanson, G., Cloninger, A., Rosing, T., and Khandelwal, A. K. (2022). Using Neural Networks to Predict Microspatial Economic Growth. American Economic Review: Insights, 4(4):491–506.

Krugman, P. (1991). Increasing Returns and Economic Geography. Journal of Political Economy. Publisher: The University of Chicago Press.

Kuznets, S. (1955). Economic Growth and Income Inequality. The American Economic Review, 45(1):1–28. Kuznets, S., Epstein, L., and Jenks, E. (1941). National Income and Its Composition, 1919-1938, Volume I.

NBER. Backup Publisher: National Bureau of Economic Research Type: Book.

Machemedze, T., Kerr, A., and Dorrington, R. (2020). South African population projection and household survey sample weight recalibration.

Matarutse, J. (2021). EThekwini Ward Level Monthly GDP Data.

McKay, A. and Perge, E. (2015). Spatial Inequality and its Implications for Growth–Poverty-Reduction Relations. In McKay, A. and Thorbecke, E., editors, Economic Growth and Poverty Reduction in Sub- Saharan Africa: Current and Emerging Issues, page 0. Oxford University Press.

Moretti, E. (2011). Local Labor Markets. In Card, D. and Ashenfelter, O., editors, Handbook of Labor Economics, volume 4, pages 1237–1313. Elsevier.

Mudau, R. (2019). Urban Population Projection Model - SCODA Open Data Portal.

NASA (2020). Gridded Population of the World (GPW), v4 | SEDAC.

National Planning Commission (2013). An integrated and inclusive rural economy. In National Development Plan 2030 - Our Future - Make it Work.

Nell, A. and Visagie, J. (2022). Spatial Tax Panel 2014-2021. Version 2.1.

Noble, M. and Wright, G. (2013). Using Indicators of Multiple Deprivation to Demonstrate the Spatial Legacy of Apartheid in South Africa. Social Indicators Research, 112(1):187–201. Publisher: Springer.

Obikili, N. (2019). The Impact of Political Competition on Economic Growth: Evidence from Municipalities in South Africa. South African Journal of Economics, 87(1):3–21. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/saje.12208.

Pretorius, L. and Clifford, C. (2019). 'One South Africa for All!' Fact-checking the DA's 2019 election manifesto.

Rebelo, A. (2020). Housing Tool.

Sacks, J. (2014). Fact check: Despite City's claims, data shows Cape Town.

Southern Africa Labour and Development Research Unit (2017). National Income Dynamics Study Wave 1-5. Technical report, Cape Town.

Statistics South Africa (2022). Mid-Year Population Estimates.

Turok, I. (2012). Urbanisation and Development in South Africa: Economic Imperatives, Spatial Distortions and Strategic Responses. ISBN: 9781843698906.

Turok, I. (2021). Urbanization, Agglomeration, and Economic Development in South Africa. In Oqubay, A., Tregenna, F., and Valodia, I., editors, The Oxford Handbook of the South African Economy, page 0. Oxford University Press.

Venables, A. J. (2016). New Economic Geography. In The New Palgrave Dictionary of Economics, pages 1–7. Palgrave Macmillan UK, London.

von Fintel, D. (2018). Long-run spatial inequality in South Africa : early settlement patterns and separate development. Studies in Economics and Econometrics, 42(2):81–102.

von Fintel, D. and Fourie, J. (2019). The great divergence in South Africa: Population and wealth dynamics over two centuries. Journal of Comparative Economics, 47(4):759–773.

von Fintel, D. and Moses, E. (2017). Migration and gender in South Africa: Following bright lights and the fortunes of others? Regional Science Policy & Practice, 9(4):251–268. _eprint: https://rsaiconnect.onlinelibrary.wiley.com/doi/pdf/ 10.1111/rsp3.12106.

Western Cape Government (2022). Municipal Economic Review and Outlook 2022/23 | Provincial Treasury. Wittenberg, M. (2017a). Measurement of earnings: Comparing South African tax and survey data. Working Paper. Accepted: 2017-10-13T12:22:56Z ISBN: 9781928281733.

Wittenberg, M. (2017b). Wages and Wage Inequality in South Africa 1994–2011: Part 1 – Wage Measurement and Trends. South African Journal of Economics, 85(2):279–297. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/saje.12148.

Wittenberg, M. (2017c). Wages and Wage Inequality in South Africa 1994–2011: Part 2 – In- equality Measurement and Trends. South African Journal of Economics, 85(2):298–318. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/saje.12147.

Appendices

Figure A.1: Population totals by metropolitan area: various sources

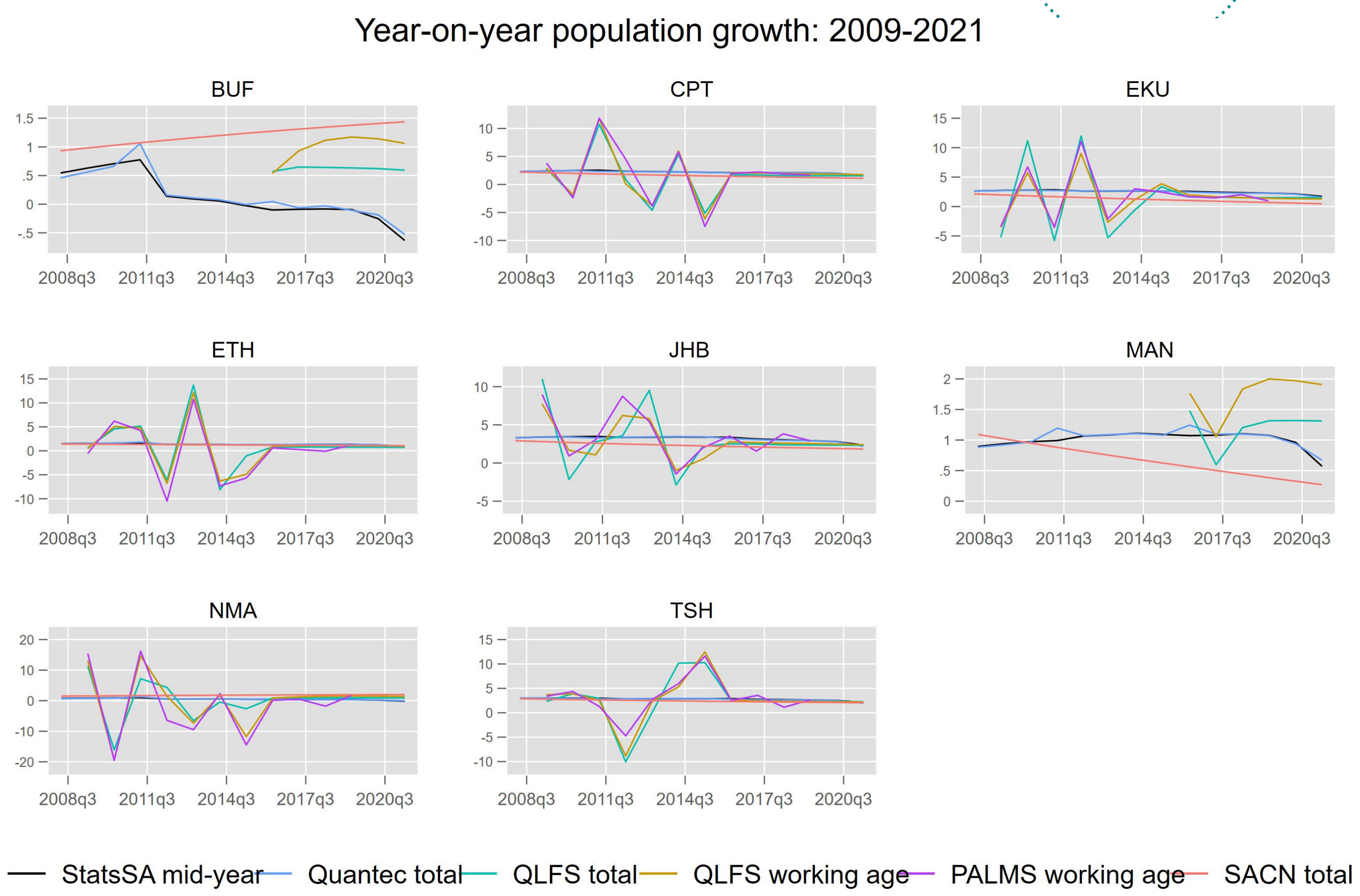


Figure A.2: Population growth by metropolitan area: various sources

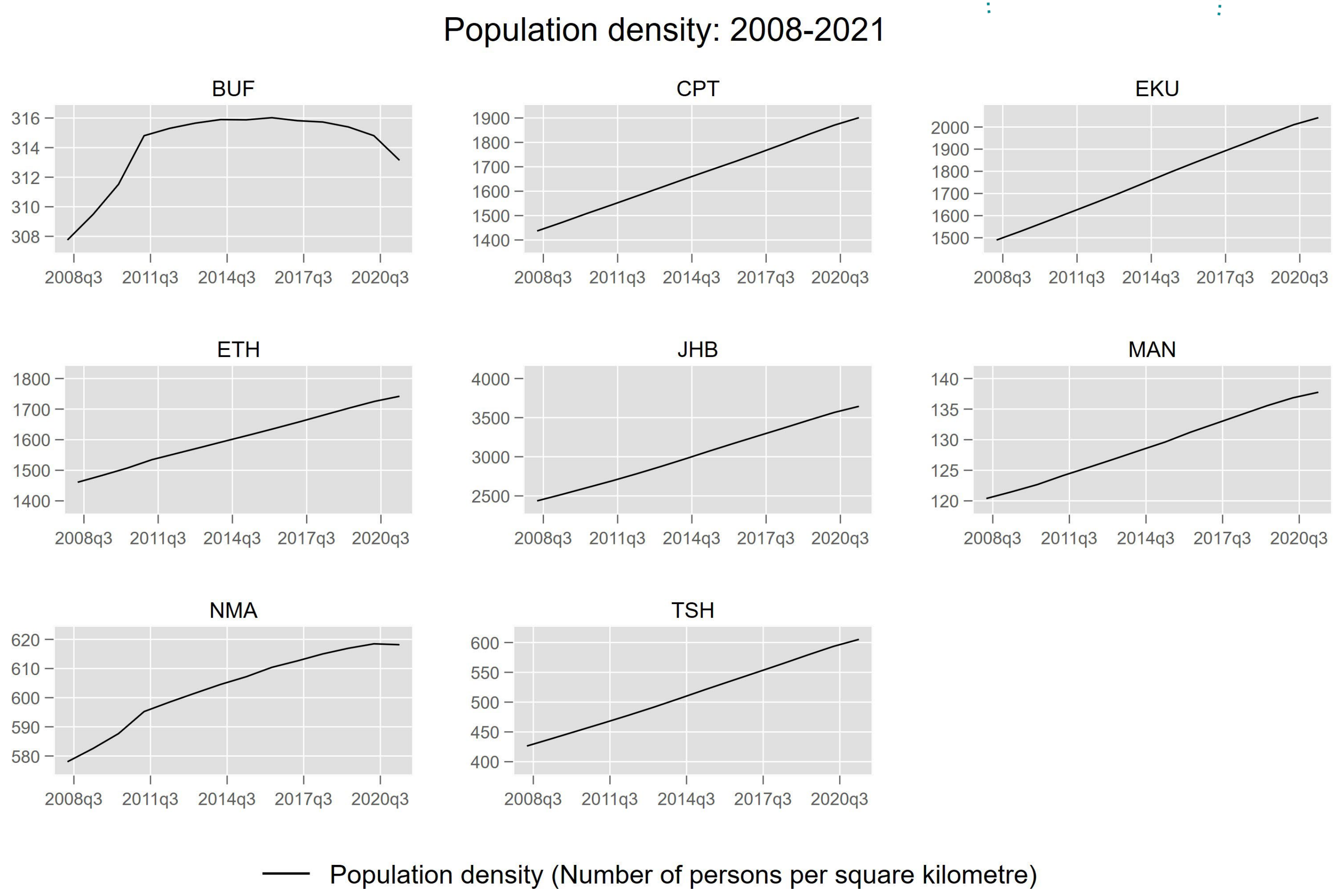


Figure B.1: Employment totals by metropolitan area: various sources

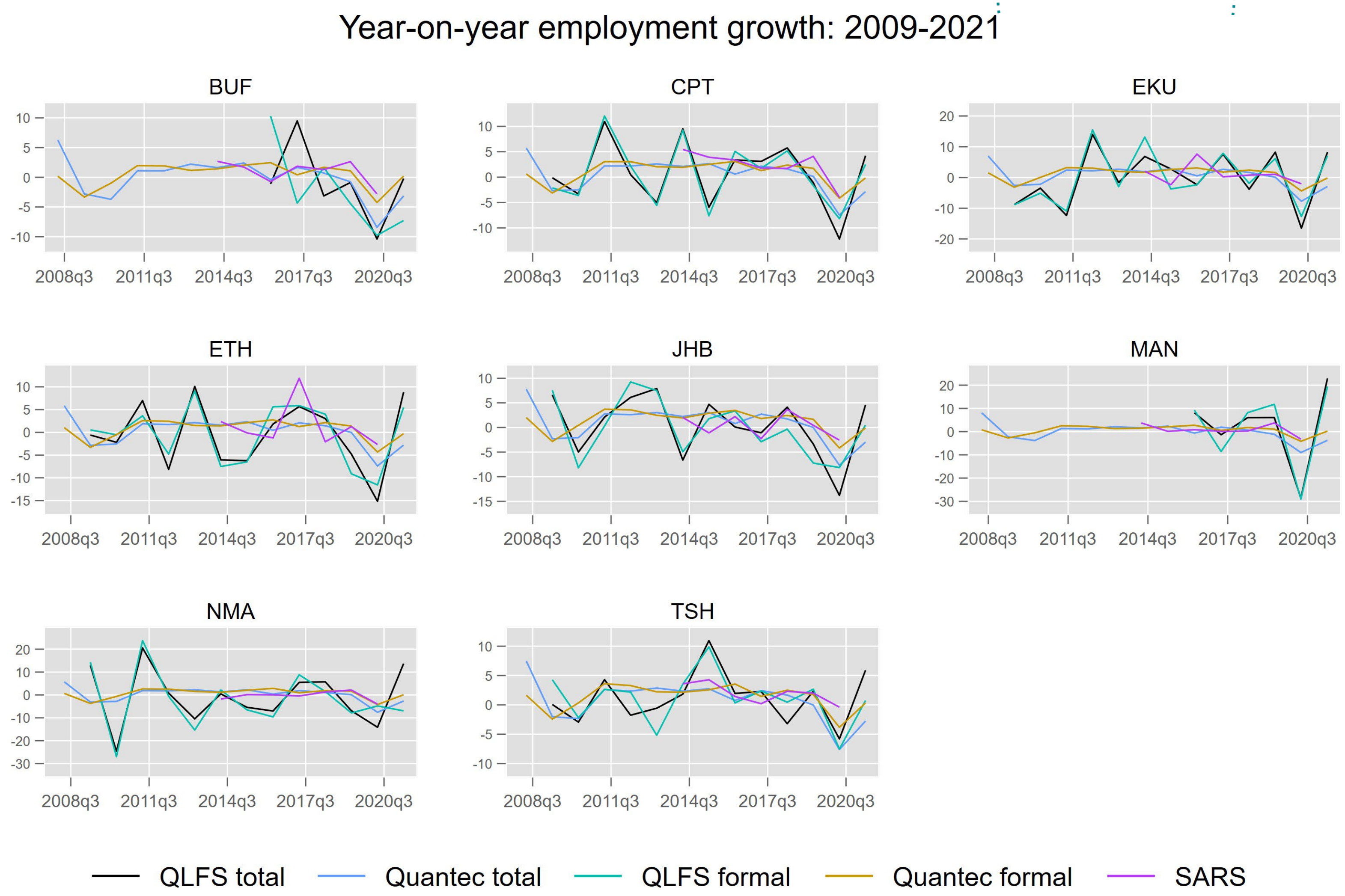


Figure B.3: Employment convergence by metropolitan area: various sources

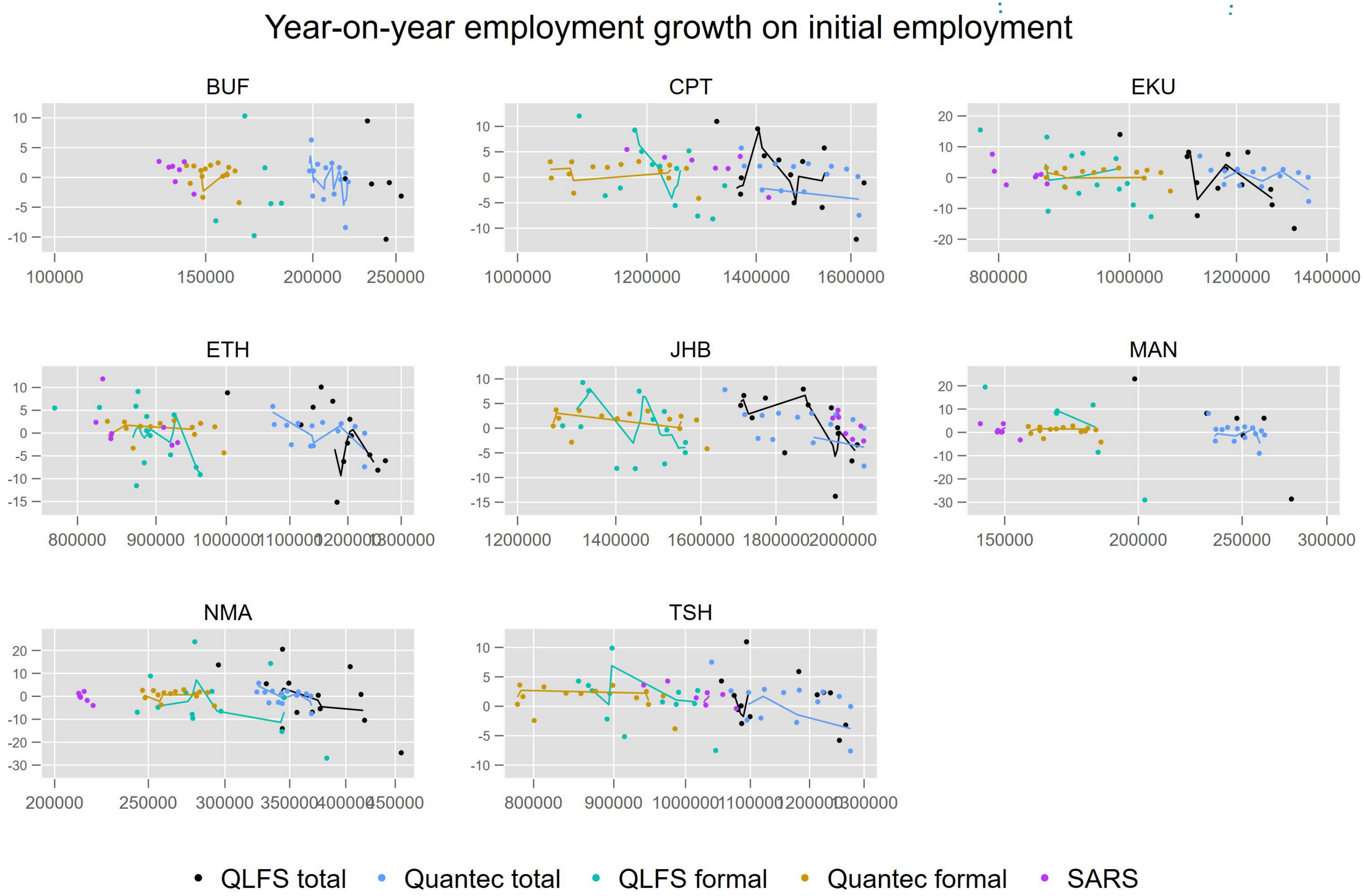


Figure B.4: Formal employment by 1 digit SIC category, source: QLFS

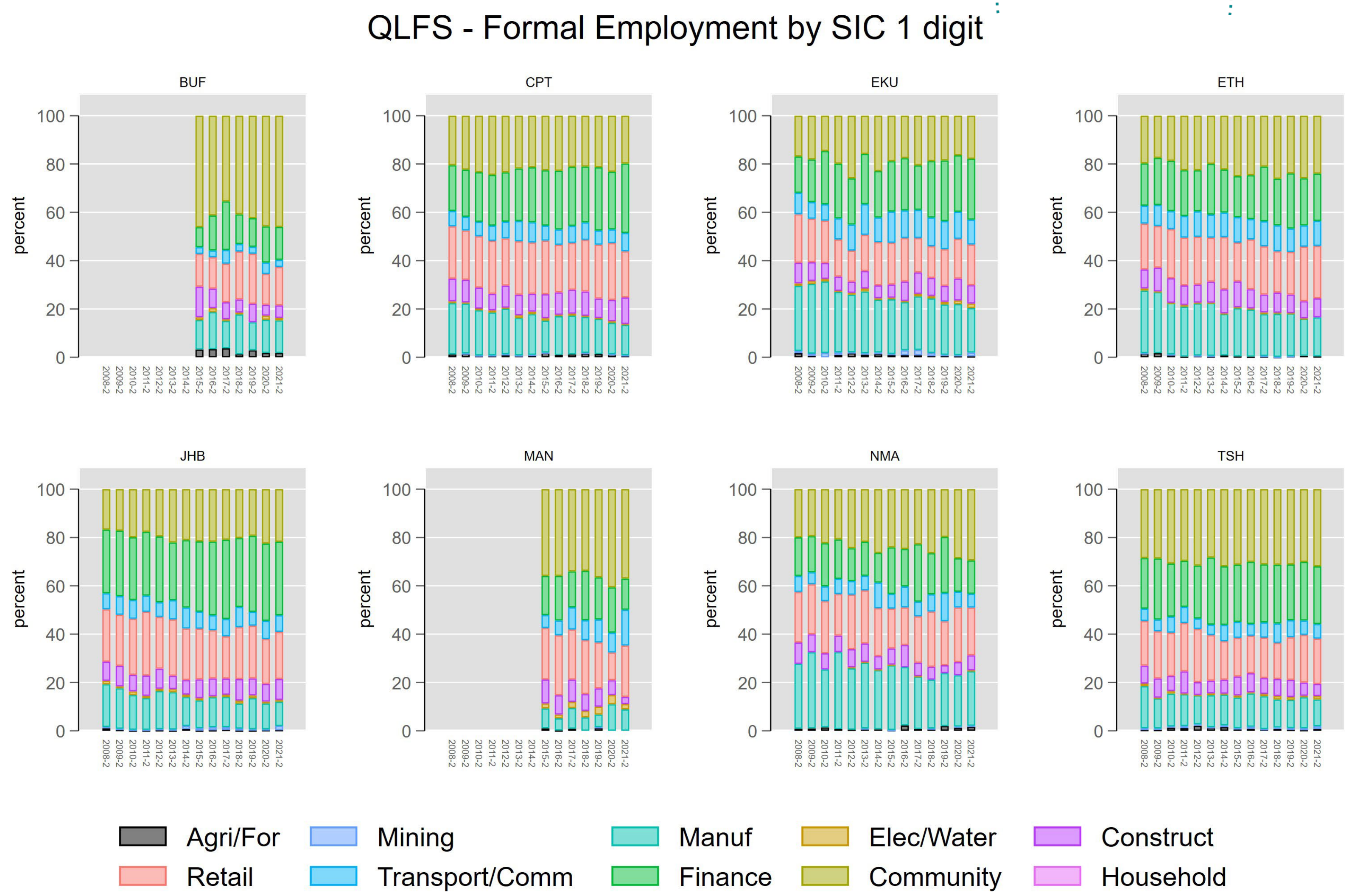


Figure B.5: Formal employment by 1 digit SIC category, source: Quantec

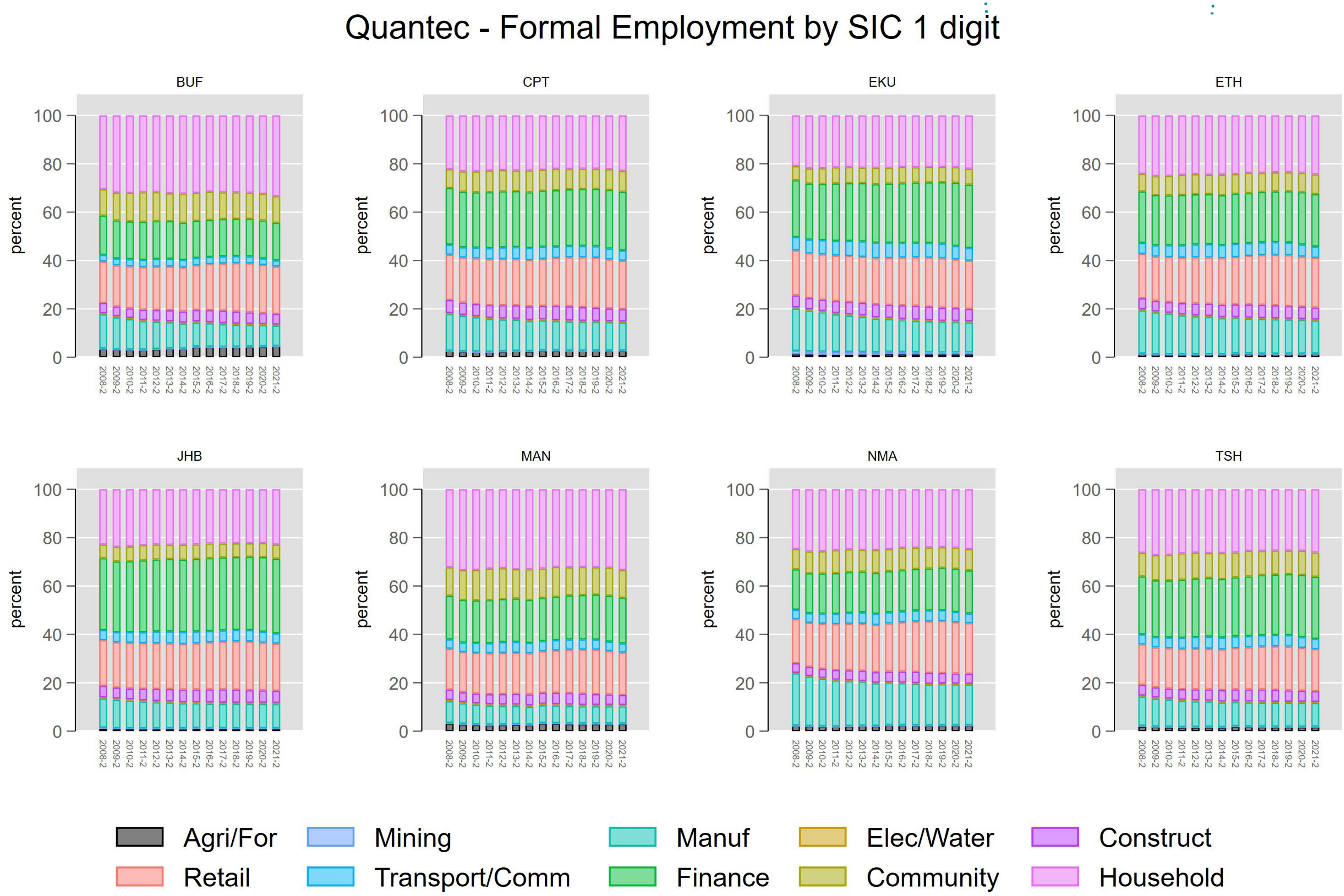


Figure B.6: Formal employment by 1 digit SIC category, source: SARS

Figure B.6: Formal employment by 1 digit SIC category, source: SARS

Figure C.1: Remuneration per worker by metropolitan area: various sources

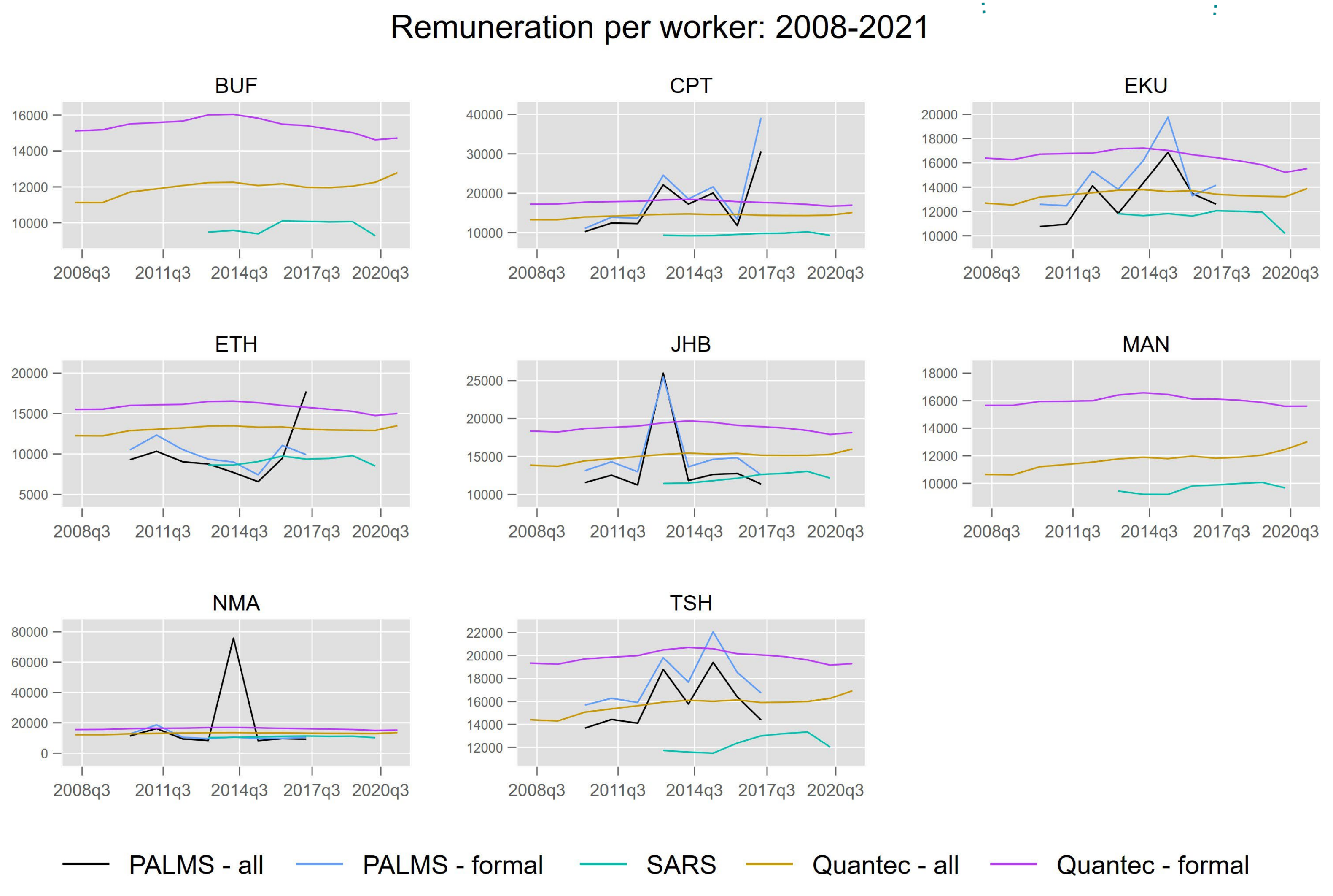


Figure C.2: Remuneration growth by metropolitan area: various sources

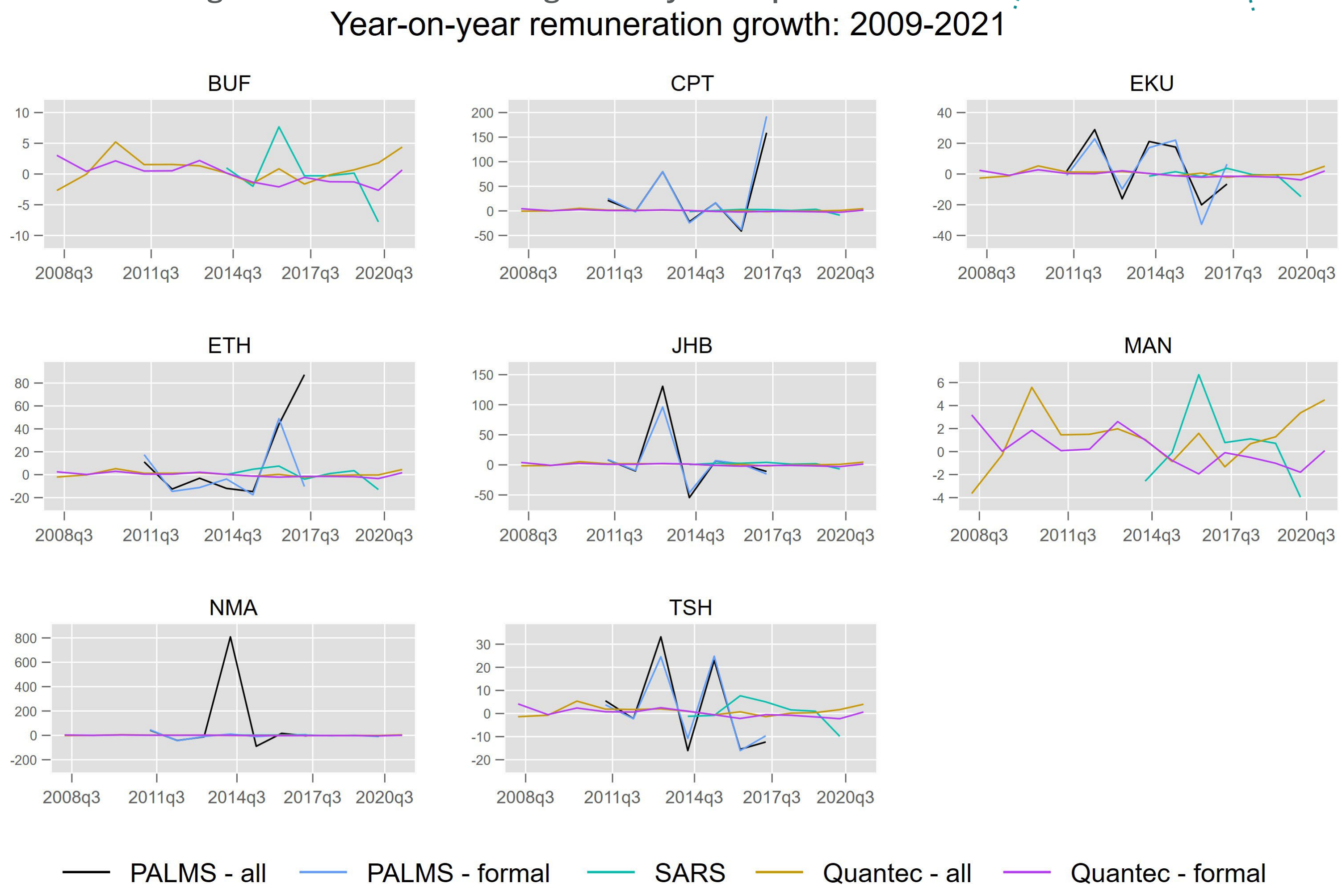


Figure C.3: Remuneration convergence by metropolitan area: various sources

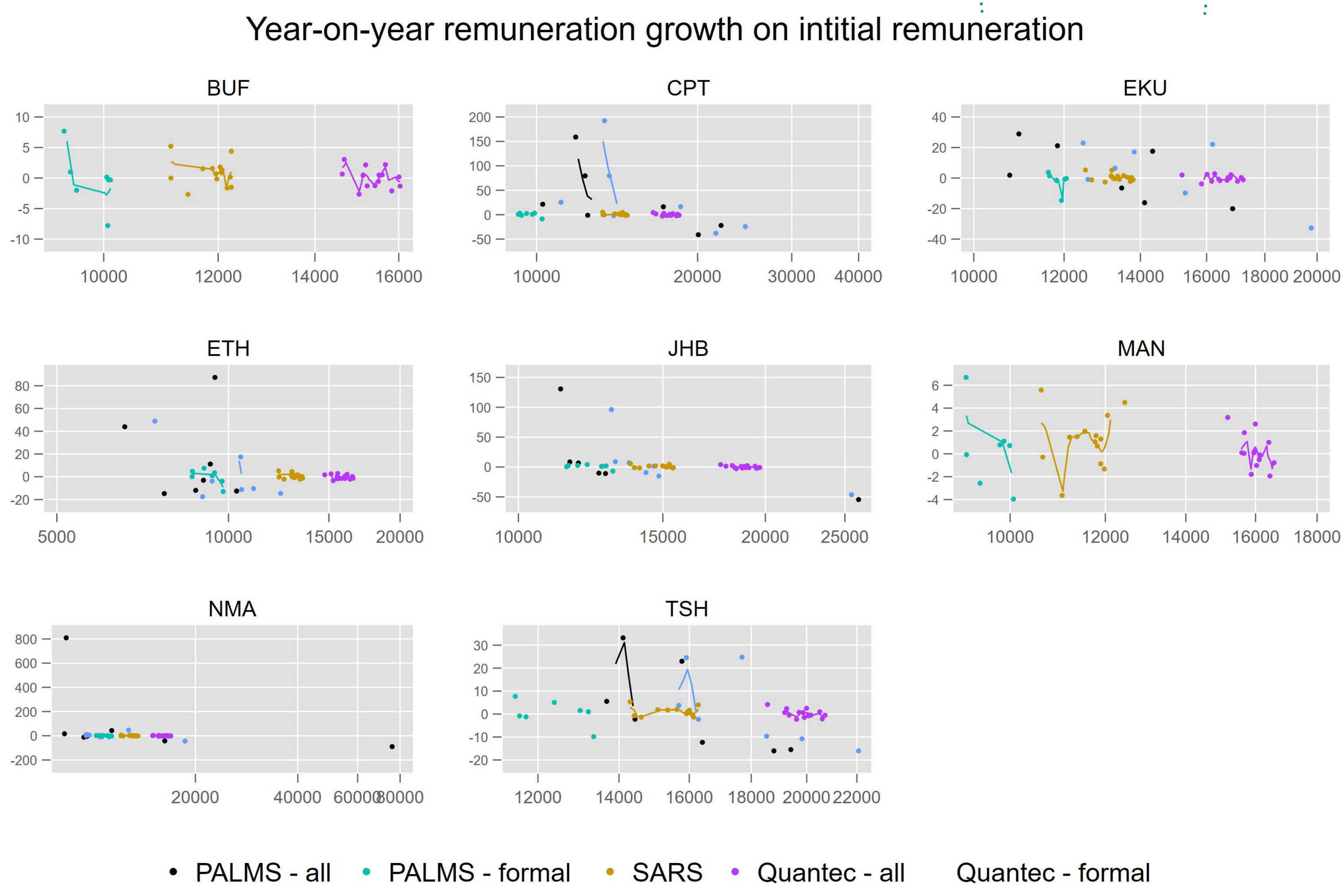


Figure C.4: SARS real median earnings

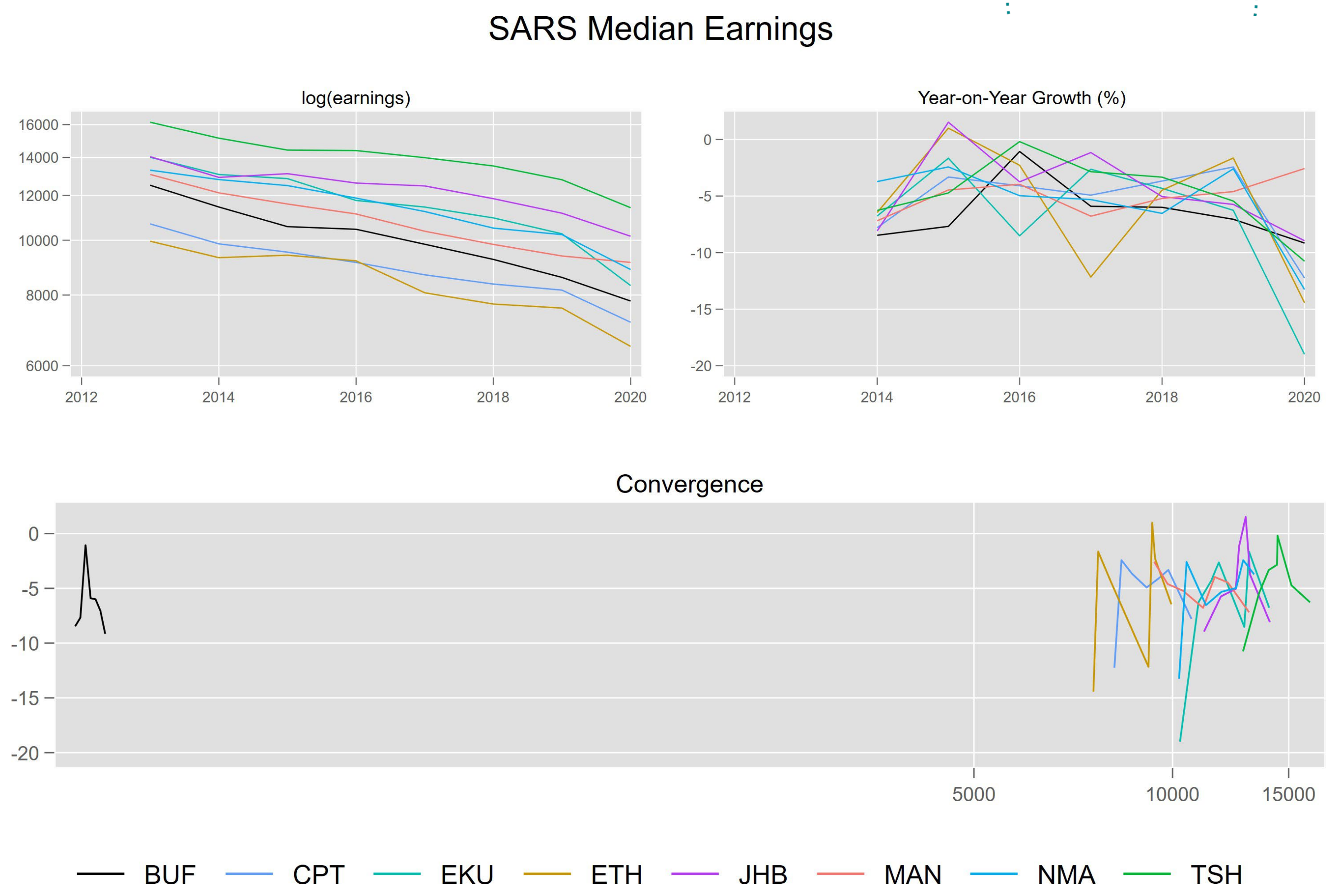


Figure D.1: Output totals by metropolitan area: various sources

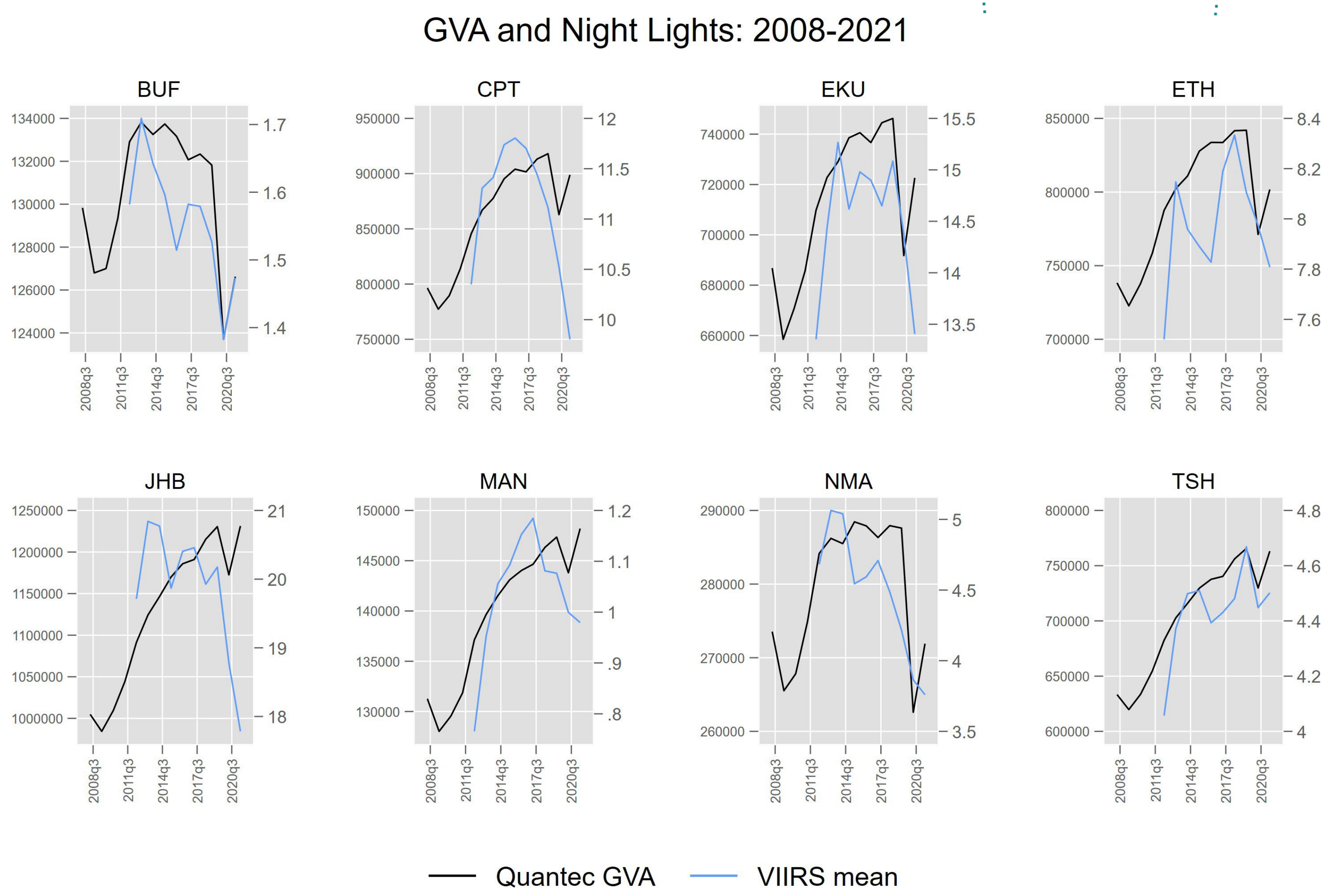


Figure D.2: Output growth by metropolitan area: various sources

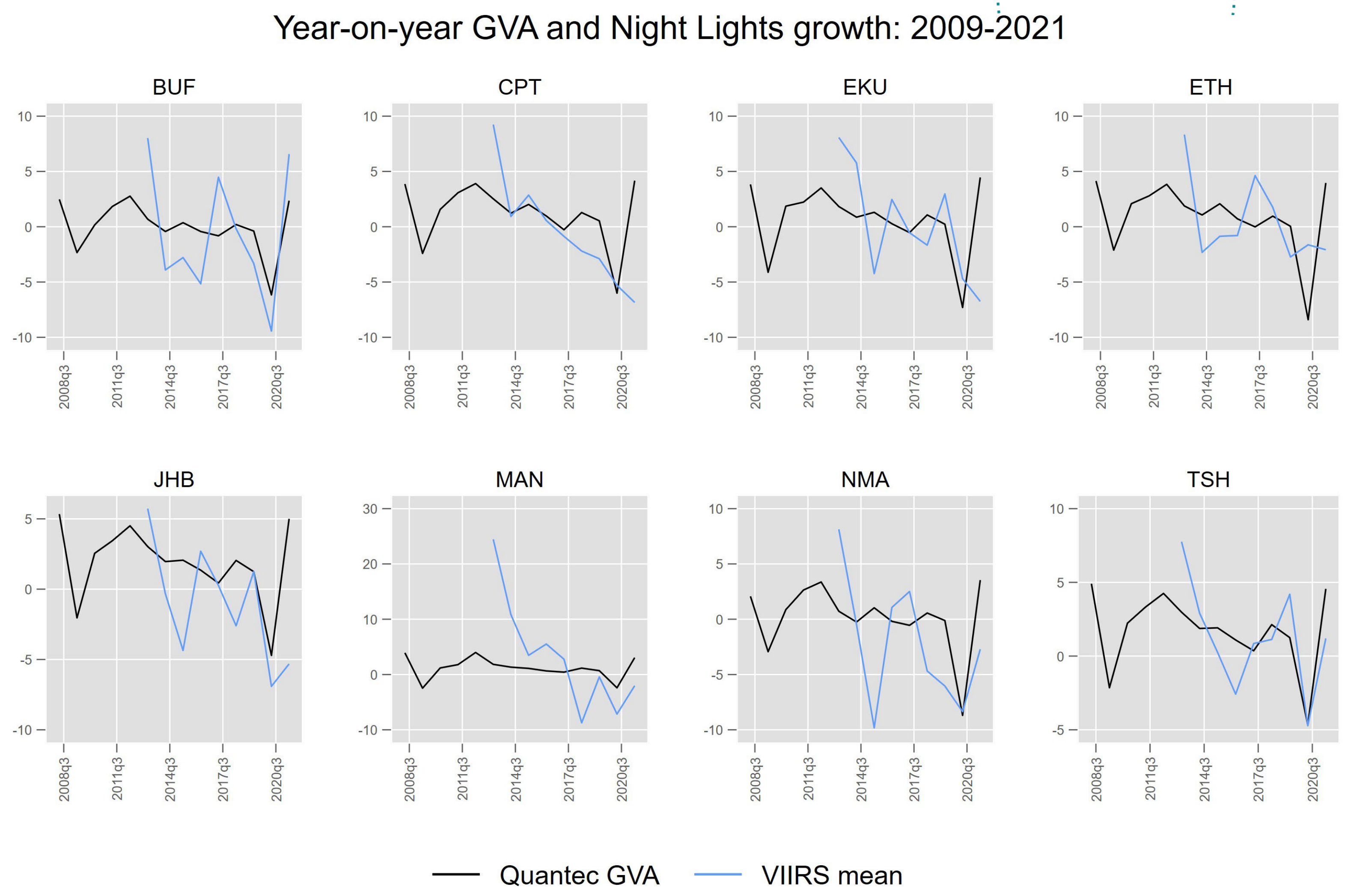


Figure D.3: Output convergence by metropolitan area: various sources

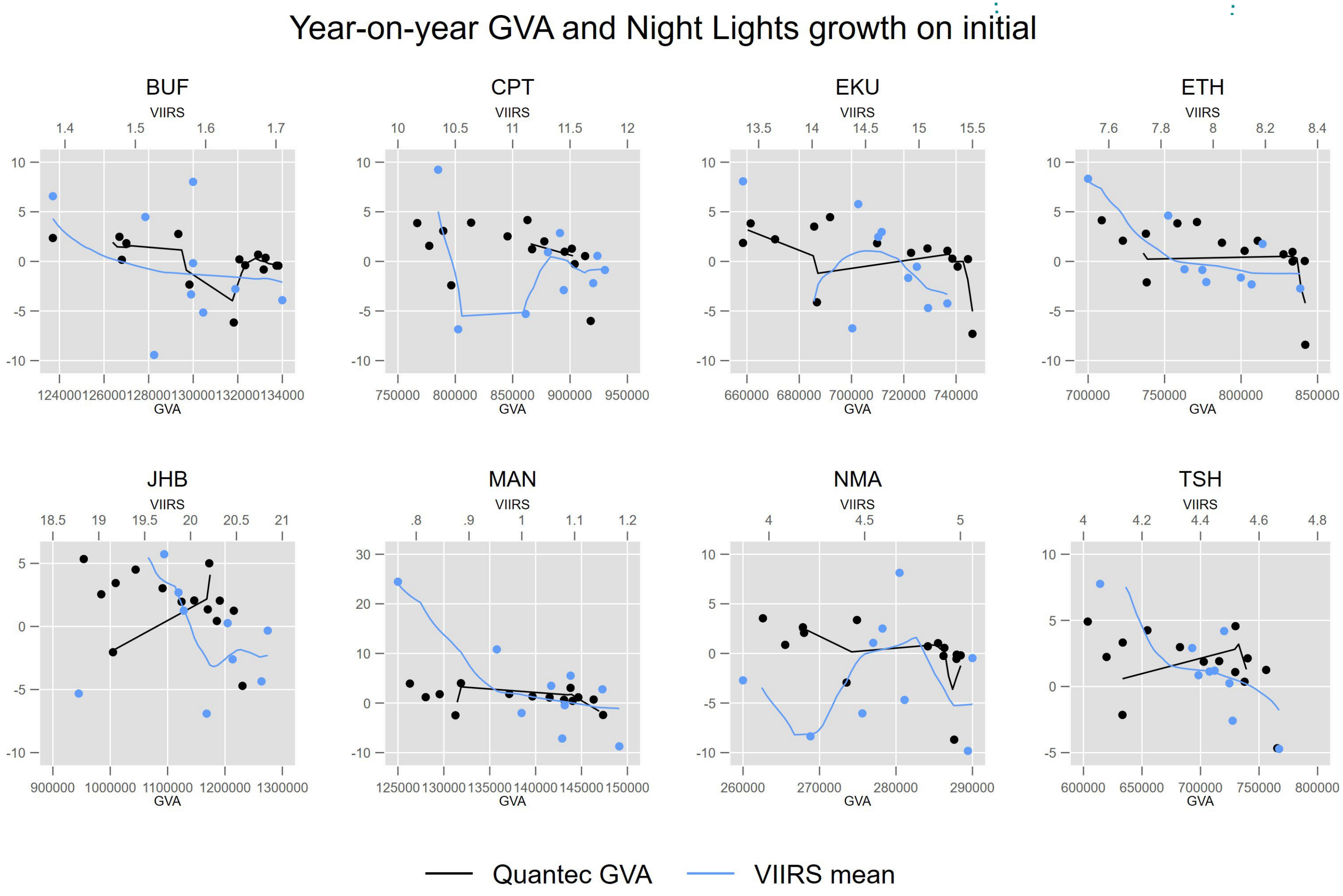
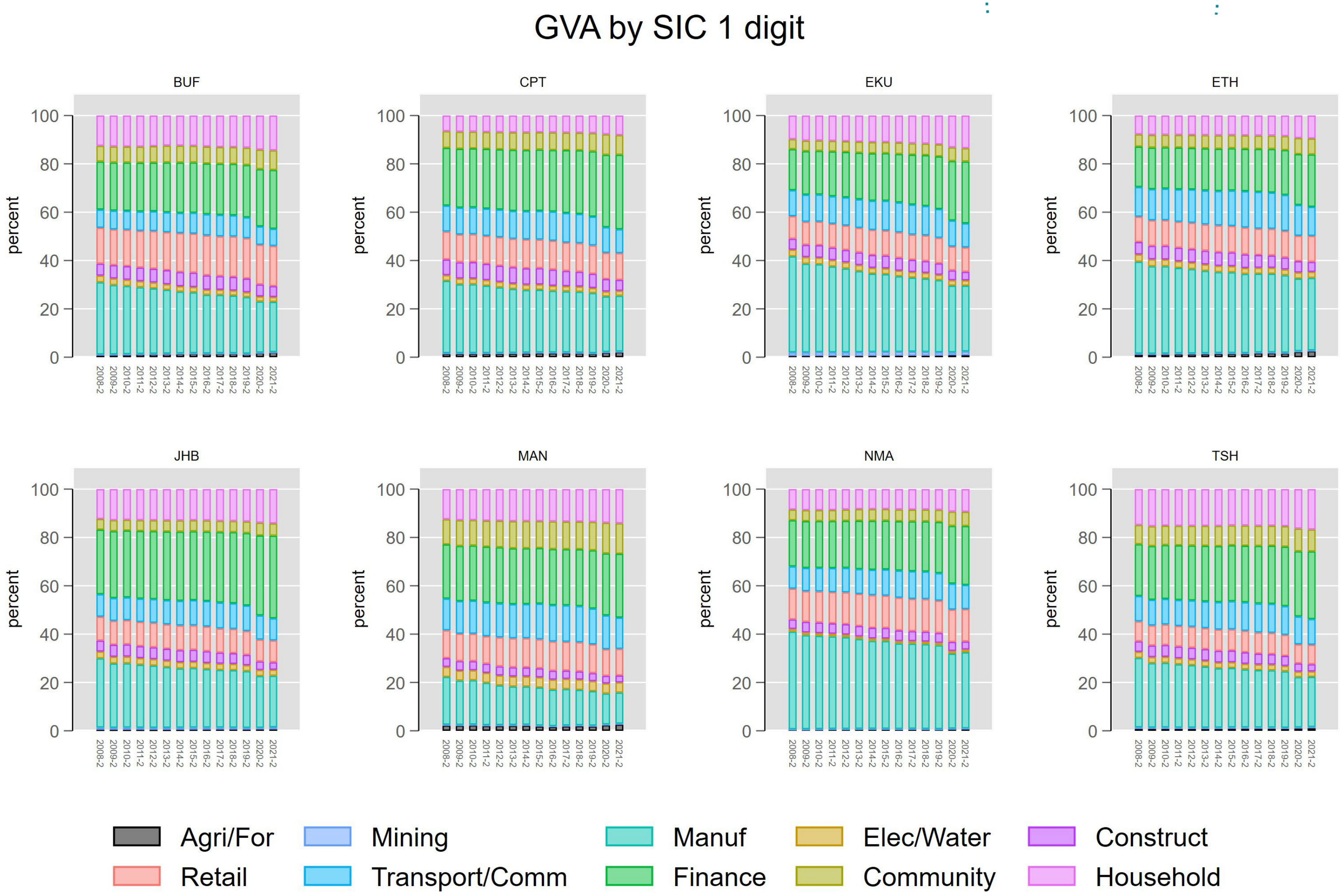


Figure D.4: GVA by SIC1 category: shares by metropolitan area (source: Quantec)



Economic Research Southern Africa (ERSA) is a platform that supports the development of economic policy by connecting economic research to national policy debate and identifying areas of future research. It has served as the premier platform for economics researchers across Southern Africa to publish their work, participate in conferences and training programmes, and contribute to the national debate on public policy, since 2004. It does this by:

- Conducting on-going research that develops and contributes to research across five broad themes.
- Sharing and promoting policy relevant economic research and code through the SAMNet Initiative.
- Stimulating discussions that contribute towards national debate, by bringing a network of economic experts to share ideas.
- Upskilling academics and students through the skills development initiative.
- Nurturing economic talent by encouraging all brains that are curious about economics to grow their knowledge and confidence in the subject.

Our network draws a broad and representative range of expert economic researchers and policy makers from a variety of academic, financial and government institutions. In this way, ERSA encourages the creation, dissemination and discussion of independent and expert economic policy-oriented research.

For more information about ERSA, please visit our website at www.econrsa.org.

Other Discussion Document Publications:

Discussion Document 08: Importance of employment programs for the youth in South Africa by Jessica Gagete-Miranda and Michelle Please

Discussion Document 09: Crime: A policy-oriented survey by Sebastian Galiani

Discussion Document 10: What luminosity data can and cannot reveal about South Africa's urban economies by Takwanisa Machemedze

Discussion Document 11: Place-based economic policies: international lessons for South Africa by Claus Rabe and Harris Selod

Discussion Document 12: Cities, productivity and jobs in South Africa: Problems and Potential by Ivan Turok and Justin Visagie